IMMUNE RESPONSE ASSOCIATED PROTEINS 10/535635

TECHNICAL FIELD

The invention relates to novel nucleic acids, immune response associated proteins encoded by these nucleic acids, and to the use of these nucleic acids and proteins in the diagnosis, treatment, and prevention of immune system, neurological, developmental, muscle, cell proliferative disorders, and disorders of lipid metabolism. The invention also relates to the assessment of the effects of exogenous compounds on the expression of nucleic acids and immune response associated proteins.

BACKGROUND OF THE INVENTION

All vertebrates have developed sophisticated and complex immune systems that provide protection from viral, bacterial, fungal and parasitic infections. Included in these systems are the processes of humoral immunity, the complement cascade and the inflammatory response (See Paul, W.E. (1993) Fundamental Immunology, Raven Press, Ltd., New York NY pp.1-20).

The cellular components of the immune system include six different types of leukocytes, or white blood cells: monocytes, lymphocytes, polymorphonuclear granulocytes (including neutrophils, eosinophils, and basophils) and plasma cells. Additionally, fragments of megakaryocytes, a seventh type of white blood cell in the bone marrow, occur in large numbers in the blood as platelets.

Leukocytes are formed from two stem cell lineages in bone marrow. The myeloid stem cell line produces granulocytes and monocytes and the lymphoid stem cell line produces lymphocytes. Lymphoid cells travel to the thymus, spleen and lymph nodes, where they mature and differentiate into lymphocytes. Leukocytes are responsible for defending the body against invading pathogens. Neutrophils and monocytes attack invading bacteria, viruses, and other pathogens and destroy them by phagocytosis. Monocytes enter tissues and differentiate into macrophages which are extremely phagocytic. Lymphocytes and plasma cells are a part of the immune system which recognizes specific foreign molecules and organisms and inactivates them, as well as signals other cells to attack the invaders.

Granulocytes and monocytes are formed and stored in the bone marrow until needed. Megakaryocytes are produced in bone marrow, where they fragment into platelets and are released into the bloodstream. The main function of platelets is to activate the blood clotting mechanism. Lymphocytes and plasma cells are produced in various lymphogenous organs, including the lymph nodes, spleen, thymus, and tonsils.

Both neutrophils and macrophages exhibit chemotaxis towards sites of inflammation. Tissue inflammation in response to pathogen invasion results in production of chemo-attractants for leukocytes, such as endotoxins or other bacterial products, prostaglandins, and products of leukocytes

or platelets.

Basophils participate in the release of the chemicals involved in the inflammatory process. The main function of basophils is secretion of these chemicals, to such a degree that they have been referred to as "unicellular endocrine glands." A distinct aspect of basophilic secretion is that the contents of granules go directly into the extracellular environment, not into vacuoles as occurs with neutrophils, eosinophils, and monocytes. Basophils have receptors for the Fc fragment of immunoglobulin E (IgE) that are not present on other leukocytes. Crosslinking of membrane IgE with anti-IgE or other ligands triggers degranulation.

Eosinophils are bi- or multi-nucleated white blood cells which contain eosinophilic granules. Their plasma membrane is characterized by Ig receptors, particularly IgG and IgE. Generally, eosinophils are stored in the bone marrow until recruited for use at a site of inflammation or invasion. They have specific functions in parasitic infections and allergic reactions, and are thought to detoxify some of the substances released by mast cells and basophils which cause inflammation. Additionally, they phagocytize antigen-antibody complexes and further help prevent the spread of inflammation.

The mononuclear phagocyte system is comprised of precursor cells in the bone marrow, monocytes in circulation, and macrophages in tissues. Macrophages are monocytes that have left the blood stream to settle in tissue. Once monocytes have migrated into tissues, they do not re-enter the bloodstream. They increase several-fold in size and transform into macrophages that are characteristic of the tissue they have entered, surviving in tissues for several months. The mononuclear phagocyte system is capable of very fast and extensive phagocytosis. A macrophage may phagocytize over 100 bacteria, digest them and extrude residues, and then survive for many more months. Macrophages are also capable of ingesting large particles, including red blood cells and malarial parasites.

Mononuclear phagocytes are essential in defending the body against invasion by foreign pathogens, particularly intracellular microorganisms such as *Mycobacterium tuberculosis*, listeria, leishmania and toxoplasma. Macrophages can also control the growth of tumorous cells, via both phagocytosis and secretion of hydrolytic enzymes. Another important function of macrophages is that of processing antigens and presenting them in a biochemically modified form to lymphocytes.

The immune system responds to invading microorganisms in two major ways: antibody production and cell mediated responses. Antibodies are immunoglobulin proteins produced by B-lymphocytes which bind to specific antigens and cause inactivation or promote destruction of the antigen by other cells. Cell-mediated immune responses involve T-lymphocytes (T cells) that react with foreign antigens on the surface of infected host cells. Depending on the type of T cell, the T cell either kills the infected cell itself, or secretes signals which activate macrophages and other cells to destroy the infected cell (Paul, *supra*).

T-lymphocytes originate in the bone marrow or liver in fetuses. Precursor cells migrate via the blood to the thymus, where they are processed to mature into T-lymphocytes. This processing is crucial because it involves positive and negative selection of T cells for those that will react with foreign antigen and not with self molecules. After processing, T cells continuously circulate in the blood and secondary lymphoid tissues, such as lymph nodes, spleen, certain epithelium-associated tissues in the gastrointestinal tract, respiratory tract and skin. When T-lymphocytes are presented with the complementary antigen, they are stimulated to proliferate and release large numbers of activated T cells into the lymph system and the blood system. These activated T cells can survive and circulate for several days. At the same time, T memory cells are created, which remain in the lymphoid tissue for months or years. Upon subsequent exposure to that specific antigen, these memory cells will respond more rapidly and with a stronger response than induced by the original antigen. This creates an "immunological memory" that can provide immunity for years.

There are two major types of T cells: cytotoxic T cells destroy infected host cells, and helper T cells activate other white blood cells via chemical signals. One class of helper cell, T_H1, activates macrophages to destroy ingested microorganisms, while another, T_H2, stimulates the production of antibodies by B cells.

Cytotoxic T cells directly attack the infected target cell. Receptors on the surface of T cells bind to antigen presented by MHC molecules on the surface of the infected cell. Once activated by binding to antigen, T cells secrete γ-interferon, a signal molecule that induces the expression of genes necessary for presenting viral (or other) antigens to cytotoxic T cells. Cytotoxic T cells kill the infected cell by stimulating programmed cell death.

Helper T cells constitute up to 75% of the total T cell population. They regulate the immune functions by producing a variety of lymphokines that act on other cells in the immune system and on bone marrow. Among these lymphokines are interleukins 2 through 6, granulocyte-monocyte colony stimulating factor, and γ-interferon.

Helper T cells are required for most B cells to respond to antigen. When an activated helper cell contacts a B cell, its centrosome and Golgi apparatus become oriented toward the B cell, aiding the directing of signal molecules, such as a transmembrane-bound protein called CD40 ligand, onto the B cell surface to interact with the CD40 transmembrane protein. Secreted signals also help B cells to proliferate and mature and, in some cases, to switch the class of antibody being produced.

B-lymphocytes (B cells) produce antibodies which react with specific antigenic proteins presented by pathogens. Once activated, B cells become filled with extensive rough endoplasmic reticulum and are known as plasma cells. As with T cells, interaction of B cells with antigen stimulates proliferation of only those B cells which produce antibody specific to that antigen. There are five classes of antibodies, known as immunoglobulins, which together comprise about 20% of

total plasma protein. Each class mediates a characteristic biological response after antigen binding. Upon activation by specific antigen B cells switch from making the membrane-bound antibody to the secreted form of that antibody.

Antibodies, or immunoglobulins, are the founding members of the immunoglobulin (Ig) superfamily and the central components of the humoral immune response. Antibodies are either expressed on the surface of B cells or secreted by B cells into the circulation. Antibodies bind and neutralize blood-borne foreign antigens. The prototypical antibody is a tetramer consisting of two identical heavy polypeptide chains (H-chains) and two identical light polypeptide chains (L-chains) interlinked by disulfide bonds. This arrangement confers the characteristic Y-shape to antibody molecules. Antibodies are classified based on their H-chain composition. The five antibody classes, IgA, IgD, IgE, IgG and IgM, are defined by the α, δ, ϵ, γ, and μ H-chain types. There are two types of L-chains, κ and λ, either of which may associate as a pair with any H-chain pair. IgG, the most common class of antibody found in the circulation, is tetrameric, while the other classes of antibodies are generally variants or multimers of this basic structure.

H-chains and L-chains each contain an N-terminal variable region and a C-terminal constant region. The constant region consists of about 110 amino acids in L-chains and about 330 or 440 amino acids in H-chains. The amino acid sequence of the constant region is nearly identical among H- or L-chains of a particular class. The variable region consists of about 110 amino acids in both H- and L-chains. However, the amino acid sequence of the variable region differs among H- or L-chains of a particular class. Within each H- or L-chain variable region are three hypervariable regions of extensive sequence diversity, each consisting of about 5 to 10 amino acids. In the antibody molecule, the H- and L-chain hypervariable regions come together to form the antigen recognition site. (Reviewed in Alberts, B. et al. (1994) Molecular Biology of the Cell, Garland Publishing, New York, NY, pp.1206-1213 and 1216-1217.)

The immune system is capable of recognizing and responding to any foreign molecule that enters the body. Therefore, the immune system must be armed with a full repertoire of antibodies against all potential antigens. Such antibody diversity is generated by somatic rearrangement of gene segments encoding variable and constant regions. These gene segments are joined together by site-specific recombination which occurs between highly conserved DNA sequences that flank each gene segment. Because there are hundreds of different gene segments, millions of unique genes can be generated combinatorially. In addition, imprecise joining of these segments and an unusually high rate of somatic mutation within these segments further contribute to the generation of a diverse antibody population.

Both H-chains and L-chains contain repeated Ig domains. For example, a typical H-chain contains four Ig domains, three of which occur within the constant region and one of which occurs

within the variable region and contributes to the formation of the antigen recognition site. Likewise, a typical L-chain contains two Ig domains, one of which occurs within the constant region and one of which occurs within the variable region. In addition, H chains such as μ have been shown to associate with other polypeptides during differentiation of the B-cell.

Antibodies can be described in terms of their two main functional domains. Antigen recognition is mediated by the Fab (antigen binding fragment) region of the antibody, while effector functions are mediated by the Fc (crystallizable fragment) region. Binding of antibody to an antigen, such as a bacterium, triggers the destruction of the antigen by phagocytic white blood cells such as macrophages and neutrophils. These cells express surface receptors that specifically bind to the antibody Fc region and allow the phagocytic cells to engulf, ingest, and degrade the antibody-bound antigen. The Fc receptors expressed by phagocytic cells are single-pass transmembrane glycoproteins of about 300 to 400 amino acids (Sears, D.W. et al. (1990) J. Immunol. 144:371-378). The extracellular portion of the Fc receptor typically contains two or three Ig domains.

Diseases which cause over- or under-abundance of any one type of leukocyte usually result in the entire immune defense system becoming involved. The most notorious autoimmune disease is AIDS (Acquired Immunodeficiency Syndrome). This disease depletes the number of helper T cells and leaves the patient susceptible to infection by microorganisms and parasites.

Another widespread medical condition attributable to the immune system is that of allergic reactions to certain antigens. Delayed reaction allergy is experienced by many genetically normal people. In the case of atopic allergies, there is a genetic origin, such that large quantities of IgE antibodies are produced. IgEs have a strong tendency to attach to mast cells and basophils, up to half million each (IgE/mast) which then rupture and release histamine, leukotrienes, eosinophil chemotactic substance, protease, neutrophil chemotactic substance, heparin, and platelet activation factors. Tissues can respond in a number of ways to these substances resulting in what are commonly known as allergic reactions: hay fever, asthma, anaphylaxis, and urticaria (hives).

Leukemias are an excess production of white blood cells, to the point where a major portion of the body's metabolic resources are directed solely at proliferation of white blood cells, leaving other tissues to starve. With lymphogenous leukemias, cancerous lymphogenous cells spread from a lymph node to other body parts. Excess T- and B-lymphocytes are produced. In myelogenous leukemias, cancerous young myelogenous cells spread from the bone marrow to other organs, especially the spleen, liver, lymph nodes and other highly vascularized regions. Usually, the extra leukemic cells released are immature, incapable of function, and undifferentiated. Occasionally, partially differentiated cells are produced, leading to classification of the disease as neutrophilic leukemia, eosinophilic leukemia, basophilic leukemia, or monocytic leukemia. Leukemias may be caused by exposure to environmental factors such as radiation or toxic chemicals or by genetic

aberration.

Leukopenia or agranulocytosis occurs when the bone marrow stops producing white blood cells. This leaves the body unprotected against foreign microorganisms, including those which normally inhabit skin, mucous membranes, and gastrointestinal tract. If all white blood cell production stops completely, infection will occur within two days and death may follow only 1 to 4 days later. Acute leukopenia can be caused by exposure to radiation or chemicals containing benzene. Occasionally, drugs such as chloramphenicol and thiouracil can suppress blood cell production by the bone marrow and initiate the onset of agranulocytosis. In cases of monoblastic leukemia, primitive monocytes in blood and bone marrow do not mature. Clinical symptoms reflect this abnormality: high lysozyme levels in blood serum, renal tubular dysfunction, and high fevers.

Impaired phagocytosis occurs in several diseases, including monocytic leukemia, systemic lupus, and granulomatous disease. In such a situation, macrophages can phagocytize normally, but the enveloped organism is not killed. There is a defect in the plasma membrane enzyme which converts oxygen to lethally reactive forms. This results in abscess formation in liver, lungs, spleen, lymph nodes, and beneath the skin.

Eosinophilia is an excess of eosinophils commonly observed in patients with allergies (hay fever, asthma), allergic reactions to drugs, rheumatoid arthritis, and cancers (Hodgkins disease, lung, and liver cancer). The mechanism for elevated levels of eosinophils in these diseases is unknown (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, Inc., New York, NY).

Host defense is further augmented by the complement system. The complement system serves as an effector system and is involved in infectious agent recognition. It can function as an independent immune network or in conjunction with other humoral immune responses. The complement system is comprised of numerous plasma and membrane proteins that act in a cascade of reaction sequences whereby one component activates the next. The result is a rapid and amplified response to infection through either an inflammatory response or increased phagocytosis.

The complement system has more than 30 protein components which can be divided into functional groupings including modified serine proteases, membrane-binding proteins, and regulators of complement activation. Activation occurs through two different pathways, the classical and the alternative. Both pathways serve to destroy infectious agents through distinct triggering mechanisms that eventually merge with the involvement of the component C3.

The classical pathway requires antibody binding to infectious agent antigens. The antibodies serve to define the target and initiate the complement system cascade, culminating in the destruction of the infectious agent. In this pathway, since the antibody guides initiation of the process, the complement system can be seen as an effector arm of the humoral immune system.

The alternative pathway of the complement system does not require the presence of pre-existing antibodies for targeting infectious agent destruction. Rather, this pathway, through low levels of an activated component, remains constantly primed and provides surveillance in the non-immune host to enable targeting and destruction of infectious agents. In this case foreign material triggers the cascade, thereby facilitating phagocytosis or lysis (Paul, *supra* pp.918-919).

Another important component of host defense is the process of inflamation. Inflammatory responses are divided into four categories on the basis of pathology and include allergic inflammation, cytotoxic antibody mediated inflammation, immune complex mediated inflammation, and monocyte mediated inflammation. Inflammation manifests as a combination of each of these forms with one predominating.

Allergic acute inflamation is observed in individuals wherein specific antigens stimulate IgE antibody production. Mast cells and basophils are subsequently activated by the attachment of antigen-IgE complexes, resulting in the release of cytoplasmic granule contents such as histamine. The products of activated mast cells can increase vascular permeability and constrict the smooth muscle of breathing passages, resulting in anaphylaxis or asthma.

Acute inflamation is also mediated by cytotoxic antibodies and can result in the destruction of tissue through the binding of complement-fixing antibodies to cells. In this case the antibodies responsible are of the IgG or IgM types and resultant clinical disorders including autoimmune hemolytic anemia and thrombocytopenia as associated with systemic lupus erythematosis.

Immune complex mediated acute inflammation involves the IgG or IgM antibody types which combine with antigen to activate the complement cascade. When such immune complexes bind to neutrophils and macrophages they activate the respiratory burst to form protein and vessel damaging agents such as hydrogen peroxide, hydroxyl radical, hypochlorous acid, and chloramines. Clinical manifestations include rheumatoid arthritis and systemic lupus erythematosus.

In chronic inflammation or delayed-type hypersensitivity, macrophages are activated and process antigen for presentation to T cells that subsequently produce lymphokines and monokines. This type of inflammatory response is likely important for defense against intracellular parasites and certain viruses. Clinical associations include granulomatous disease, tuberculosis, leprosy, and sarcoidosis (Paul, *supra* pp.1017-1018).

Most cell surface and soluble molecules that mediate functions such as recognition, adhesion or binding have evolved from a common evolutionary precursor (i.e., these proteins have structural homology). A number of molecules outside the immune system that have similar functions are also derived from this same evolutionary precursor. These molecules are classified as members of the immunoglobulin (Ig) superfamily. The criteria for a protein to be a member of the Ig superfamily is to have one or more Ig domains, which are regions of 70-110 amino acid residues in length

homologous to either Ig variable-like (V) or Ig constant-like (C) domains. Members of the Ig superfamily include antibodies (Ab), T cell receptors (TCRs), class I and II major histocompatibility (MHC) proteins, CD2, CD3, CD4, CD8, poly-Ig receptors, Fc receptors, neural cell-adhesion molecule (NCAM) and platelet-derived growth factor receptor (PDGFR).

Ig domains (V and C) are regions of conserved amino acid residues that give a polypeptide a globular tertiary structure called an immunoglobulin (or antibody) fold, which consists of two approximately parallel layers of β-sheets. Conserved cysteine residues form an intrachain disulfide-bonded loop, 55-75 amino acid residues in length, which connects the two layers of the β-sheets. Each β-sheet has three or four anti-parallel β-strands of 5-10 amino acid residues. Hydrophobic and hydrophilic interactions of amino acid residues within the β-strands stabilize the Ig fold (hydrophobic on inward facing amino acid residues and hydrophilic on the amino acid residues in the outward facing portion of the strands). A V domain consists of a longer polypeptide than a C domain, with an additional pair of β-strands in the Ig fold.

A consistent feature of Ig superfamily genes is that each sequence of an Ig domain is encoded by a single exon. It is possible that the superfamily evolved from a gene coding for a single Ig domain involved in mediating cell-cell interactions. New members of the superfamily then arose by exon and gene duplications. Modern Ig superfamily proteins contain different numbers of V and/or C domains. Another evolutionary feature of this superfamily is the ability to undergo DNA rearrangements, a unique feature retained by the antigen receptor members of the family.

Many members of the Ig superfamily are integral plasma membrane proteins with extracellular Ig domains. The hydrophobic amino acid residues of their transmembrane domains and their cytoplasmic tails are very diverse, with little or no homology among Ig family members or to known signal-transducing structures. There are exceptions to this general superfamily description. For example, the cytoplasmic tail of PDGFR has tyrosine kinase activity. In addition Thy-1 is a glycoprotein found on thymocytes and T cells. This protein has no cytoplasmic tail, but is instead attached to the plasma membrane by a covalent glycophosphatidylinositol linkage.

Another common feature of many Ig superfamily proteins is the interactions between Ig domains which are essential for the function of these molecules. Interactions between Ig domains of a multimeric protein can be either homophilic or heterophilic (i.e., between the same or different Ig domains). Antibodies are multimeric proteins which have both homophilic and heterophilic interactions between Ig domains. Pairing of constant regions of heavy chains forms the Fc region of an antibody and pairing of variable regions of light and heavy chains form the antigen binding site of an antibody. Heterophilic interactions also occur between Ig domains of different molecules. These interactions provide adhesion between cells for significant cell-cell interactions in the immune system and in the developing and mature nervous system. (Reviewed in Abbas, A.K. et al. (1991)

Cellular and Molecular Immunology, W.B. Saunders Company, Philadelphia, PA, pp.142-145.)

Sushi domains, also known as complement control protein (CCP) modules, or short consensus repeats (SCR), are found in a wide variety of complement and adhesion proteins. CD21 (also called C3d receptor, CR2, Epstein Barr virus receptor or EBV-R) is the receptor for EBV and for C3d, C3dg and iC3b. Complement components may activate B cells through CD21. CD21 is part of a large signal-transduction complex that also involves CD19, CD81, and Leu13. Some of the proteins in this group are responsible for the molecular basis of the blood group antigens, surface markers on the outside of the red blood cell membrane. Most of these markers are proteins, but some are carbohydrates attached to lipids or proteins (for a review see Reid, M.E. and C. Lomas-Francis (1977) The Blood Group Antigen FactsBook Academic Press, San Diego, CA). Complement decay-accelerating factor (Antigen CD55) belongs to the Cromer blood group system and is associated with Cr(a), Dr(a), Es(a), Tc(a/b/c), Wd(a), WES(a/b), IFC and UMC antigens. Complement receptor type 1 (C3b/C4b receptor) (Antigen CD35) belongs to the Knops blood group system and is associated with Kn(a/b), McC(a), Sl(a) and Yk(a) antigens.

Human leukocyte-specific transcript 1 (LST1) is a small protein that modulates immune responses and cellular morphogenesis. LST1 is expressed at high levels in dendritic cells. A DNA-binding site and interaction of multiple regulatory elements may be involved in mediating the expression of the various forms of LST1 mRNA (Yu, X. and Weissman, S.M. (2000) J. Biol. Chem. 275:34597-34608).

Spalpha is a member of the scavenger receptor cysteine-rich (SRCR) family of proteins. Spalpha is expressed only in lymphoid tissues, where it is implicated in monocyte activity (Gebe, J.A. (1997) J. Biol. Chem. 272:6151-6158).

A family of metalloproteases, the ADAMs (for A Disintegrin and Metalloprotease Domain), has been shown to play a role in the immune system (Yamamoto, S. et al.(1999) Immunol. Today 20:278-84). These proteins share with their close relatives the adamalysins, snake venom metalloproteases (SVMPs). ADAMs combine features of both cell surface adhesion molecules and proteases, containing a prodomain, a protease domain, a disintegrin domain, a cysteine rich domain, an epidermal growth factor repeat, a transmembrane domain, and a cytoplasmic tail. The first three domains listed above are also found in the SVMPs. The ADAMs possess four potential functions: proteolysis, adhesion, signaling and fusion. The ADAMs share the metzincin zinc binding sequence and are inhibited by some MMP antagonists such as TIMP-1.

ADAMs are implicated in such processes as sperm-egg binding and fusion, myoblast fusion, and protein-ectodomain processing or shedding of cytokines, cytokine receptors, adhesion proteins and other extracellular protein domains (Schlöndorff, J. and C.P. Blobel (1999) J. Cell. Sci. 112:3603-3617). The Kuzbanian protein cleaves a substrate in the NOTCH pathway (possibly

NOTCH itself), activating the program for lateral inhibition in *Drosophila* neural development. Two ADAMs, TACE (ADAM 17) and ADAM 10, are proposed to have analogous roles in the processing of amyloid precursor protein in the brain (Schlöndorff and Blobel, *supra*). TACE has also been identified as the TNF activating enzyme (Black, R.A. et al. (1997) Nature 385:729-733). TNF is a pleiotropic cytokine that is important in mobilizing host defenses in response to infection or trauma, but can cause severe damage in excess and is often overproduced in autoimmune disease. TACE cleaves membrane-bound pro-TNF to release a soluble form. Other ADAMs may be involved in a similar type of processing of other membrane-bound molecules.

Expression profiling

Microarrays are analytical tools used in bioanalysis. A microarray has a plurality of molecules spatially distributed over, and stably associated with, the surface of a solid support. Microarrays of polypeptides, polynucleotides, and/or antibodies have been developed and find use in a variety of applications, such as gene sequencing, monitoring gene expression, gene mapping, bacterial identification, drug discovery, and combinatorial chemistry.

One area in particular in which microarrays find use is in gene expression analysis. Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

Breast Cancer

More than 180,000 new cases of breast cancer are diagnosed each year, and the mortality rate for breast cancer approaches 10% of all deaths in females between the ages of 45-54 (Gish, K. (1999) AWIS Magazine 28:7-10). However the survival rate based on early diagnosis of localized breast cancer is extremely high (97%), compared with the advanced stage of the disease in which the tumor has spread beyond the breast (22%). Current procedures for clinical breast examination are lacking in sensitivity and specificity, and efforts are underway to develop comprehensive gene expression profiles for breast cancer that may be used in conjunction with conventional screening methods to improve diagnosis and prognosis of this disease (Perou, C.M. et al. (2000) Nature 406:747-752).

Mutations in two genes, BRCA1 and BRCA2, are known to greatly predispose a woman to breast cancer and may be passed on from parents to children (Gish, *supra*). However, this type of hereditary breast cancer accounts for only about 5% to 9% of breast cancers, while the vast majority

of breast cancer is due to non-inherited mutations that occur in breast epithelial cells.

The relationship between expression of epidermal growth factor (EGF) and its receptor, EGFR, to human mammary carcinoma has been particularly well studied. (See Khazaie, K. et al. (1993) Cancer and Metastasis Rev. 12:255-274, and references cited therein for a review of this area.) Overexpression of EGFR, particularly coupled with down-regulation of the estrogen receptor, is a marker of poor prognosis in breast cancer patients. In addition, EGFR expression in breast tumor metastases is frequently elevated relative to the primary tumor, suggesting that EGFR is involved in tumor progression and metastasis. This is supported by accumulating evidence that EGF has effects on cell functions related to metastatic potential, such as cell motility, chemotaxis, secretion and differentiation. Changes in expression of other members of the erbB receptor family, of which EGFR is one, have also been implicated in breast cancer. The abundance of erbB receptors, such as HER-2/neu, HER-3, and HER-4, and their ligands in breast cancer points to their functional importance in the pathogenesis of the disease, and may therefore provide targets for therapy of the disease (Bacus, S.S. et al. (1994) Am. J. Clin. Pathol. 102:S13-S24). Other known markers of breast cancer include a human secreted frizzled protein mRNA that is downregulated in breast tumors; the matrix G1a protein which is overexpressed in human breast carcinoma cells; Drg1 or RTP, a gene whose expression is diminished in colon, breast, and prostate tumors; maspin, a tumor suppressor gene downregulated in invasive breast carcinomas; and CaN19, a member of the S100 protein family, all of which are down-regulated in mammary carcinoma cells relative to normal mammary epithelial cells (Zhou, Z. et al. (1998) Int. J. Cancer 78:95-99; Chen, L. et al. (1990) Oncogene 5:1391-1395; Ulrix, W. et al (1999) FEBS Lett 455:23-26; Sager, R. et al. (1996) Curr. Top. Microbiol. Immunol. 213:51-64; and Lee, S.W. et al. (1992) Proc. Natl. Acad. Sci. USA 89:2504-2508).

Cell lines derived from human mammary epithelial cells at various stages of breast cancer provide a useful model to study the process of malignant transformation and tumor progression as it has been shown that these cell lines retain many of the properties of their parental tumors for lengthy culture periods (Wistuba, I.I. et al. (1998) Clin. Cancer Res. 4:2931-2938). Such a model is particularly useful for comparing phenotypic and molecular characteristics of human mammary epithelial cells at various stages of malignant transformation.

Colon cancer

While soft tissue sarcomas are relatively rare, more than 50% of new patients diagnosed with the disease will die from it. The molecular pathways leading to the development of sarcomas are relatively unknown, due to the rarity of the disease and variation in pathology. Colon cancer evolves through a multi-step process whereby pre-malignant colonocytes undergo a relatively defined sequence of events leading to tumor formation. Several factors participate in the process of

tumor progression and malignant transformation including genetic factors, mutations, and selection.

To understand the nature of gene alterations in colorectal cancer, a number of studies have focused on the inherited syndromes. Familial adenomatous polyposis (FAP), is caused by mutations in the adenomatous polyposis coli gene (APC), resulting in truncated or inactive forms of the protein. This tumor suppressor gene has been mapped to chromosome 5q. Hereditary nonpolyposis colorectal cancer (HNPCC) is caused by mutations in mis-match repair genes. Although hereditary colon cancer syndromes occur in a small percentage of the population and most colorectal cancers are considered sporadic, knowledge from studies of the hereditary syndromes can be generally applied. For instance, somatic mutations in APC occur in at least 80% of sporadic colon tumors. APC mutations are thought to be the initiating event in the disease. Other mutations occur subsequently. Approximately 50% of colorectal cancers contain activating mutations in ras, while 85% contain inactivating mutations in p53. Changes in all of these genes lead to gene expression changes in colon cancer.

Lung cancer

Lung cancer is the leading cause of cancer death for men and the second leading cause of cancer death for women in the U.S. Lung cancers are divided into four histopathologically distinct groups. Three groups (squamous cell carcinoma, adenocarcinoma, and large cell carcinoma) are classified as non-small cell lung cancers (NSCLCs). The fourth group of cancers is referred to as small cell lung cancer (SCLC). Deletions on chromosome 3 are common in this disease and are thought to indicate the presence of a tumor suppressor gene in this region. Activating mutations in K-ras are commonly found in lung cancer and are the basis of one of the mouse models for the disease.

Obesity

The potential application of gene expression profiling is particularly relevant to improving diagnosis, prognosis, and treatment of disease. For example, both the levels and sequences expressed in tissues from subjects with obesity or type II diabetes may be compared with the levels and sequences expressed in normal tissue.

The primary function of adipose tissue is the ability to store and release fat during periods of feeding and fasting. White adipose tissue is the major energy reserve in periods of excessive energy use, and its reserve is mobilized during energy deprivation. Understanding how various molecules regulate adiposity and energy balance in physiological and pathophysiological situations may lead to the development of novel therapeutics for human obesity. Adipose tissue is one of the primary target tissues for insulin. Adipogenesis and insulin resistance are linked in type II diabetes, non-insulin dependent diabetes mellitus (NIDDM). Most patients with type II diabetes are obese and obesity, in turn, causes insulin resistance. Cytologically the conversion of a preadipocytes into mature

adipocytes is characterized by deposition of fat droplets around the nuclei. The conversion process in vivo can be induced by thiazolidinediones and other PPARγ agonists (Adams et al. (1997) J Clin Invest 100:3149-3153) which also lead to increased sensitivity to insulin and reduced plasma glucose and blood pressure.

Pickup and Crook (1998; Diabetologia 41:1241-8) have suggested that NIDDM may result from the inability of an individual with hypersensitive acute-phase immune response to carry out normal cell signaling and repair. Steps in this process are highly correlated with long-term lifestyle and environment and include: 1) high glucose stimulation of insulin and cytokine production, 2) influence of various cytokines on tissue remodeling during adipocyte differentiation and their affect on signaling pathways, and 3) occurrence of tissue damage when cytokines continue to be produced, extracellular matrix components (ECM) are not recycled, and homeostasis is not timely restored. Many cytokines and the receptors with which they interact are implicated in this process. These cytokines include tumor necrosis factor, connective tissue growth factor, transforming growth factor-beta, interleukin (IL)-13 and their receptors. Tumor necrosis factor contributes to insulin resistance by inhibiting insulin-stimulated tyrosine phosphorylation of the insulin receptor. This, in turn, prevents the insulin receptor from participating in normal signaling processes (Skolnik and Marcusohn (1996) Cytokine Growth Factor Rev 7:161-173; Hotamisligil (1999) J Intern med 245:621-625). Connective tissue growth factor mediates the buildup of mesenglial matrix (Murphy et al. (2000) J Biol Chem 274:5830-5834). Transforming growth factor-beta mediates the buildup of mesenglial matrix of the kidney and affects vascular function through its interaction with the inositol trisphosphate receptor, a key intracellular calcium channel (Sharma and McGowan (2000) Cytokine Growth Factor Rev 11:115-123).

IL-13 and IL-4 are immuno-regulatory cytokines which share many overlapping biological properties. They both promote growth of B-cells (McKenzie et al. (1993) Proc Natl Acad Sci 90:3735-3739), induce expression of germ line Cε transcripts, and direct naive B cells to switch to the synthesis of IgE and IgG4 (Punnomen et al. (1993) Proc Natl Acad Sci 90:3730-3734). Similarly, different isoforms of the IL-13 and IL-4 receptors interact to form four types of IL-13 receptor complexes. In some instances, IL-13 utilizes a receptor complex composed of the IL-4 receptor-α chain (Rα) and the IL-13Rα. Although the specific role of each chain in IL-13 signaling is unclear, Ba/F3 cells transfected with IL-13Rα1 display a mitogenic response to IL-13, but cells transfected with mouse IL-13Rα2 do not. In addition, a soluble IL-13Rα2/Fc fusion protein blocks the mitogenic response to IL-13 (Donaldson et al. (1998) J Immunol 161:2317-2324). This suggests that IL-13Rα2 could serve as a dominant negative inhibitor or decoy receptor for IL-13. However, in colonic carcinoma cell lines, the receptor complex displayed growth inhibition which was associated with tyrosine phosphorylation of insulin receptor substrate-1. It is evident that more

research is needed to establish 1) which isoforms of the receptor complex promote cell growth and which inhibit cell growth and 2) whether this varies by cell or tissue type.

The majority of research in adipocyte biology to date has been done using transformed mouse preadipocyte cell lines. The culture condition which stimulates mouse preadipocyte differentiation is different from that for inducing human primary preadipocyte differentiation. In addition, primary cells are diploid and may therefore reflect the *in vivo* context better than aneuploid cell lines. Understanding the gene expression profile during adipogenesis in humans will lead to an understanding of the fundamental mechanism of adiposity regulation. Furthermore, through comparing the gene expression profiles of adipogenesis between donor with normal weight and donor with obesity, identification of crucial genes, potential drug targets for obesity and type II diabetes, will be possible.

Thiazolidinediones (TZDs) act as agonists for the peroxisome-proliferator-activated receptor gamma (PPARγ), a member of the nuclear hormone receptor superfamily. TZDs reduce hyperglycemia, hyperinsulinemia, and hypertension, in part by promoting glucose metabolism and inhibiting gluconeogenesis. Roles for PPARγ and its agonists have been demonstrated in a wide range of pathological conditions including diabetes, obesity, hypertension, atherosclerosis, polycystic ovarian syndrome, and cancers such as breast, prostate, liposarcoma, and colon cancer.

The mechanism by which TZDs and other PPARγ agonists enhance insulin sensitivity is not fully understood, but may involve the ability of PPARγ to promote adipogenesis. When ectopically expressed in cultured preadipocytes, PPARγ is a potent inducer of adipocyte differentiation. TZDs, in combination with insulin and other factors, can also enhance differentiation of human preadipocytes in culture (Adams et al. (1997) J. Clin. Invest. 100:3149-3153). The relative potency of different TZDs in promoting adipogenesis *in vitro* is proportional to both their insulin sensitizing effects *in vivo*, and their ability to bind and activate PPARγ *in vitro*. Interestingly, adipocytes derived from omental adipose depots are refractory to the effects of TZDs. It has therefore been suggested that the insulin sensitizing effects of TZDs may result from their ability to promote adipogenesis in subcutaneous adipose depots (Adams et al., *supra*). Further, dominant negative mutations in the PPARγ gene have been identified in two non-obese subjects with severe insulin resistance, hypertension, and overt non-insulin dependent diabetes mellitus (NIDDM) (Barroso et al. (1998) Nature 402:880-883).

NIDDM is the most common form of diabetes mellitus, a chronic metabolic disease that affects 143 million people worldwide. NIDDM is characterized by abnormal glucose and lipid metabolism that results from a combination of peripheral insulin resistance and defective insulin secretion. NIDDM has a complex, progressive etiology and a high degree of heritability. Numerous complications of diabetes including heart disease, stroke, renal failure, retinopathy, and peripheral

neuropathy contribute to the high rate of morbidity and mortality.

At the molecular level, PPARγ functions as a ligand activated transcription factor. In the presence of ligand, PPARγ forms a heterodimer with the retinoid X receptor (RXR) which then activates transcription of target genes containing one or more copies of a PPARγ response element (PPRE). Many genes important in lipid storage and metabolism contain PPREs and have been identified as PPARγ targets, including PEPCK, aP2, LPL, ACS, and FAT-P (Auwerx, J. (1999) Diabetologia 42:1033-1049). Multiple ligands for PPARγ have been identified. These include a variety of fatty acid metabolites; synthetic drugs belonging to the TZD class, such as Pioglitazone and Rosiglitazone (BRL49653); and certain non-glitazone tyrosine analogs such as GI262570 and GW1929. The prostaglandin derivative 15-dPGJ2 is a potent endogenous ligand for PPARγ.

Expression of PPARγ is very high in adipose but barely detectable in skeletal muscle, the primary site for insulin stimulated glucose disposal in the body. PPARγ is also moderately expressed in large intestine, kidney, liver, vascular smooth muscle, hematopoietic cells, and macrophages. The high expression of PPARγ in adipose tissue suggests that the insulin sensitizing effects of TZDs may result from alterations in the expression of one or more PPARγ regulated genes in adipose tissue. Identification of PPARγ target genes will contribute to better drug design and the development of novel therapeutic strategies for diabetes, obesity, and other conditions.

Systematic attempts to identify PPARγ target genes have been made in several rodent models of obesity and diabetes (Suzuki et al. (2000) Jpn. J. Pharmacol. 84:113-123; Way et al. (2001) Endocrinology 142:1269-1277). However, a serious drawback of the rodent gene expression studies is that significant differences exist between human and rodent models of adipogenesis, diabetes, and obesity (Taylor (1999) Cell 97:9-12; Gregoire et al. (1998) Physiol. Reviews 78:783-809). Therefore, an unbiased approach to identifying TZD regulated genes in primary cultures of human tissues is necessary to fully elucidate the molecular basis for diseases associated with PPARγ activity.

Ovarian Cancer

Ovarian cancer is the leading cause of death from a gynecologic cancer. The majority of ovarian cancers are derived from epithelial cells, and 70% of patients with epithelial ovarian cancers present with late-stage disease. As a result, the long-term survival rate for this disease is very low. Identification of early-stage markers for ovarian cancer would significantly increase the survival rate. Genetic variations involved in ovarian cancer development include mutation of p53 and microsatellite instability. Gene expression patterns likely vary when normal ovary is compared to ovarian tumors.

Promonocytes

Leukocytes comprise lymphocytes, granulocytes, and monocytes. Lymphocytes include T-

and B-cells, which specifically recognize and respond to foreign pathogens. T-cells fight viral infections and activate other leukocytes, while B-cells secrete antibodies that neutralize bacteria and other microbes. Granulocytes and monocytes are primarily migratory, phagocytic cells that exit the bloodstream to fight infection in tissues. Monocytes, which are derived from immature promonocytes, further differentiate into macrophages that engulf and digest microorganisms and damaged or dead cells. Monocytes and macrophages modulate the immune response by secreting signaling molecules such as growth factors and cytokines. Tumor necrosis factor-α (TNF-α), for example, is a macrophage-secreted protein with anti-tumor and anti-viral activity. In addition, monocytes and macrophages are recruited to sites of infection and inflammation by signaling proteins secreted by other leukocytes. The differentiation of the monocyte blood cell lineage can be studied *in vitro* using cultured cell lines. For example, THP-1 is a human promonocyte cell line that can be activated by treatment with both phorbol ester such as phorbol myristate acetate (PMA), and lipopolysaccharide (LPS). PMA is a broad activator of the protein kinase C-dependent pathways.

Monocytes are involved in the initiation and maintenance of inflammatory immune responses. The outer membrane of gram-negative bacteria expresses lipopolysaccharide (LPS) complexes called endotoxins. Toxicity is associated with the lipid component (Lipid A) of LPS, and immunogenicity is associated with the polysaccharide components of LPS. LPS elicits a variety of inflammatory responses, and because it activates complement by the alternative (properdin) pathway, it is often part of the pathology of gram-negative bacterial infections. For the most part, endotoxins remain associated with the cell wall until the bacteria disintegrate. LPS released into the bloodstream by lysing gram-negative bacteria is first bound by certain plasma proteins identified as LPS-binding proteins. The LPS-binding protein complex interacts with CD14 receptors on monocytes, macrophages, B cells, and other types of receptors on endothelial cells. Activation of human B cells with LPS results in mitogenesis as well as immunoglobulin synthesis. In monocytes and macrophages three types of events are triggered during their interaction with LPS: 1) production of cytokines, including IL-1, IL-6, IL-8, TNF-α , and platelet-activating factor, which stimulate production of prostaglandins and leukotrienes that mediate inflammation and septic shock; 2) activation of the complement cascade; and 3) activation of the coagulation cascade.

Prostate cancer

Prostate cancer is a common malignancy in men over the age of 50, and the incidence increases with age. In the US, there are approximately 132,000 newly diagnosed cases of prostate cancer and more than 33,000 deaths from the disorder each year.

Once cancer cells arise in the prostate, they are stimulated by testosterone to a more rapid growth. Thus, removal of the testes can indirectly reduce both rapid growth and metastasis of the cancer. Over 95 percent of prostatic cancers are adenocarcinomas which originate in the prostatic

acini. The remaining 5 percent are divided between squamous cell and transitional cell carcinomas, both of which arise in the prostatic ducts or other parts of the prostate gland.

As with most tumors, prostate cancer develops through a multistage progression ultimately resulting in an aggressive tumor phenotype. The initial step in tumor progression involves the hyperproliferation of normal luminal and/or basal epithelial cells. Androgen responsive cells become hyperplastic and evolve into early-stage tumors. Although early-stage tumors are often androgen sensitive and respond to androgen ablation, a population of androgen independent cells evolve from the hyperplastic population. These cells represent a more advanced form of prostate tumor that may become invasive and potentially become metastatic to the bone, brain, or lung. A variety of genes may be differentially expressed during tumor progression. For example, loss of heterozygosity (LOH) is frequently observed on chromosome 8p in prostate cancer. Fluorescence *in situ* hybridization (FISH) revealed a deletion for at least 1 locus on 8p in 29 (69%) tumors, with a significantly higher frequency of the deletion on 8p21.2-p21.1 in advanced prostate cancer than in localized prostate cancer, implying that deletions on 8p22-p21.3 play an important role in tumor differentiation, while 8p21.2-p21.1 deletion plays a role in progression of prostate cancer (Oba, K. et al. (2001) Cancer Genet. Cytogenet. 124: 20-26).

A primary diagnostic marker for prostate cancer is prostate specific antigen (PSA). PSA is a tissue-specific serine protease almost exclusively produced by prostatic epithelial cells. The quantity of PSA correlates with the number and volume of the prostatic epithelial cells, and consequently, the levels of PSA are an excellent indicator of abnormal prostate growth. Men with prostate cancer exhibit an early linear increase in PSA levels followed by an exponential increase prior to diagnosis. However, since PSA levels are also influenced by factors such as inflammation, androgen and other growth factors, some scientists maintain that changes in PSA levels are not useful in detecting individual cases of prostate cancer.

Current areas of cancer research provide additional prospects for markers as well as potential therapeutic targets for prostate cancer. Several growth factors have been shown to play a critical role in tumor development, growth, and progression. The growth factors Epidermal Growth Factor (EGF), Fibroblast Growth Factor (FGF), and Tumor Growth Factor alpha (TGFα) are important in the growth of normal as well as hyperproliferative prostate epithelial cells, particularly at early stages of tumor development and progression, and affect signaling pathways in these cells in various ways (Lin, J. et al. (1999) Cancer Res. 59:2891-2897; Putz, T. et al. (1999) Cancer Res. 59:227-233). The TGF-β family of growth factors are generally expressed at increased levels in human cancers and the high expression levels in many cases correlates with advanced stages of malignancy and poor survival (Gold, L.I. (1999) Crit. Rev. Oncog. 10:303-360). Finally, there are human cell lines representing both the androgen-dependent stage of prostate cancer (LNCap) as well as the androgen-

independent, hormone refractory stage of the disease (PC3 and DU-145) that have proved useful in studying gene expression patterns associated with the progression of prostate cancer, and the effects of cell treatments on these expressed genes (Chung, T.D. (1999) Prostate 15:199-207).

There is a need in the art for new compositions, including nucleic acids and proteins, for the diagnosis, prevention, and treatment of immune system, neurological, developmental, muscle, cell proliferative disorders, and disorders of lipid metabolism.

SUMMARY OF THE INVENTION

Various embodiments of the invention provide purified polypeptides, immune response associated proteins, referred to collectively as 'IRAP' and individually as 'IRAP-1,' 'IRAP-2,' 'IRAP-3,' 'IRAP-4,' 'IRAP-5,' 'IRAP-6,' 'IRAP-7,' 'IRAP-8,' 'IRAP-9,' 'IRAP-10,' 'IRAP-11,' 'IRAP-12,' 'IRAP-13,' 'IRAP-14,' 'IRAP-15,' 'IRAP-16,' 'IRAP-17,' 'IRAP-18,' 'IRAP-19,' 'IRAP-20,' 'IRAP-21,' 'IRAP-22,' 'IRAP-23,' 'IRAP-24,' 'IRAP-25,' 'IRAP-26,' 'IRAP-27,' 'IRAP-28,' 'IRAP-29,' 'IRAP-30,' 'IRAP-31,' and 'IRAP-32' and methods for using these proteins and their encoding polynucleotides for the detection, diagnosis, and treatment of diseases and medical conditions. Embodiments also provide methods for utilizing the purified immune response associated proteins and/or their encoding polynucleotides for facilitating the drug discovery process, including determination of efficacy, dosage, toxicity, and pharmacology. Related embodiments provide methods for utilizing the purified immune response associated proteins and/or their encoding polynucleotides for investigating the pathogenesis of diseases and medical conditions.

An embodiment provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-32. Another embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID NO:1-32.

Still another embodiment provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, c) a biologically active fragment of a

polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-32. In another embodiment, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-32. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:33-64.

Still another embodiment provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-32. Another embodiment provides a cell transformed with the recombinant polynucleotide. Yet another embodiment provides a transgenic organism comprising the recombinant polynucleotide.

Another embodiment provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-32. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Yet another embodiment provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-32.

Still yet another embodiment provides an isolated polynucleotide selected from the group

consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:33-64, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:33-64, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In other embodiments, the polynucleotide can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:33-64, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:33-64, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex. In a related embodiment, the method can include detecting the amount of the hybridization complex. In still other embodiments, the probe can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Still yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:33-64, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:33-64, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof. In a related embodiment, the method can include detecting the amount of the amplified target polynucleotide or fragment thereof.

Another embodiment provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, b) a polypeptide comprising a

naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and a pharmaceutically acceptable excipient. In one embodiment, the composition can comprise an amino acid sequence selected from the group consisting of SEQ ID NO:1-32. Other embodiments provide a method of treating a disease or condition associated with decreased or abnormal expression of functional IRAP, comprising administering to a patient in need of such treatment the composition.

Yet another embodiment provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-32. The method comprises a) contacting a sample comprising the polypeptide with a compound, and b) detecting agonist activity in the sample. Another embodiment provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with decreased expression of functional IRAP, comprising administering to a patient in need of such treatment the composition.

Still yet another embodiment provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-32. The method comprises a) contacting a sample comprising the polypeptide with a compound, and b) detecting antagonist activity in the sample. Another embodiment provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with overexpression of functional IRAP, comprising administering to a patient in need of such treatment the composition.

Another embodiment provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-32. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

Yet another embodiment provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-32. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

Still yet another embodiment provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:33-64, the method comprising a) contacting a sample comprising the target polynucleotide with a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

Another embodiment provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20

contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:33-64, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:33-64, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:33-64, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:33-64, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide can comprise a fragment of a polynucleotide selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for full length polynucleotide and polypeptide embodiments of the invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptide embodiments of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide embodiments, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide embodiments, along with selected fragments of the polynucleotides.

Table 5 shows representative cDNA libraries for polynucleotide embodiments.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze polynucleotides and polypeptides, along with applicable descriptions, references, and threshold parameters.

Table 8 shows single nucleotide polymorphisms found in polynucleotide sequences of the invention, along with allele frequencies in different human populations.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleic acids, and methods are described, it is understood that embodiments of the invention are not limited to the particular machines, instruments, materials, and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the invention.

As used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with various embodiments of the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"IRAP" refers to the amino acid sequences of substantially purified IRAP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of IRAP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of IRAP either by directly interacting with IRAP or by acting on components of the biological pathway in which IRAP participates.

An "allelic variant" is an alternative form of the gene encoding IRAP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or

many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding IRAP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as IRAP or a polypeptide with at least one functional characteristic of IRAP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding IRAP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide encoding IRAP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent IRAP. Deliberate amino acid substitutions may be made on the basis of one or more similarities in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of IRAP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" can refer to an oligopeptide, a peptide, a polypeptide, or a protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid. Amplification may be carried out using polymerase chain reaction (PCR) technologies or other nucleic acid amplification technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of IRAP. Antagonists may include proteins such as antibodies, anticalins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of IRAP either by directly interacting with IRAP or by acting on components of the biological pathway in which IRAP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments

thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind IRAP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker (Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13).

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense"

(coding) strand of a polynucleotide having a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic IRAP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide" and a "composition comprising a given polypeptide" can refer to any composition containing the given polynucleotide or polypeptide. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotides encoding IRAP or fragments of IRAP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (Accelrys, Burlington MA) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of

the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of IRAP or a polynucleotide encoding IRAP which can be identical in sequence to, but shorter in length than, the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from about 5 to about 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:33-64 can comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:33-64, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:33-64 can be employed in one or more embodiments of methods of the invention, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:33-64 from related polynucleotides. The precise length of a fragment of SEQ ID NO:33-64 and the region of SEQ ID NO:33-64 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-32 is encoded by a fragment of SEQ ID NO:33-64. A fragment of SEQ ID NO:1-32 can comprise a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-32. For example, a fragment of SEQ ID NO:1-32 can be used as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-32. The precise length of a fragment of SEQ ID NO:1-32 and the region of SEQ ID NO:1-32 to which the fragment corresponds can be determined based on the intended purpose for the fragment using one or more analytical methods described herein or otherwise known in the art.

A "full length" polynucleotide is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, alternatively, sequence identity, between two

or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of identical nucleotide matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using one or more computer algorithms or programs known in the art or described herein. For example, percent identity can be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989; CABIOS 5:151-153) and in Higgins, D.G. et al. (1992; CABIOS 8:189-191). For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms which can be used is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of identical residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide. The phrases "percent similarity" and "% similarity," as applied to polypeptide sequences, refer to the percentage of residue matches, including identical residue matches and conservative substitutions, between at least two polypeptide sequences aligned using a standardized algorithm. In contrast, conservative substitutions are not included in the calculation of percent identity between polypeptide sequences.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of

the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY, ch. 9).

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acids by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid present in solution and another nucleic acid immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or polynucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of IRAP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of IRAP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, antibodies, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, antibody, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of IRAP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of IRAP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an IRAP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of IRAP.

"Probe" refers to nucleic acids encoding IRAP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acids. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90,

100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in, for example, Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY), Ausubel, F.M. et al. (1999; Short Protocols in Molecular Biology, 4th ed., John Wiley & Sons, New York NY), and Innis, M. et al. (1990; PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a nucleic acid that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of

sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook and Russell (*supra*). The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA molecule, is composed of the same linear sequence of nucleotides as the reference DNA molecule with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing IRAP, nucleic acids encoding IRAP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60%

free, preferably at least about 75% free, and most preferably at least about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In another embodiment, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook and Russell (*supra*).

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotides that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity or sequence similarity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity or sequence similarity over a certain defined length of one of the polypeptides.

THE INVENTION

Various embodiments of the invention include new human immune response associated proteins (IRAP), the polynucleotides encoding IRAP, and the use of these compositions for the diagnosis, treatment, or prevention of immune system, neurological, developmental, muscle, cell proliferative disorders, and disorders of lipid metabolism.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide embodiments of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is

denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to polypeptide embodiments of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows amino acid residues comprising signature sequences, domains, motifs, potential phosphorylation sites, and potential glycosylation sites. Column 5 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are immune response associated proteins. For example, SEQ ID NO:4 is 99% identical, from residue M1 to residue M391, and 100% identical from residue V392 to residue K450, to human bactericidal permeability increasing protein (BPI) precursor (GenBank ID g179529) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.7e-236, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:4 also has homology to proteins that are localized to the plasma membrane, have lipopolysaccharide (LPS)-binding function, and are bactericidal/permeability-increasing proteins, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:4 also contains BPI/LBP/CETP domains, and LBP/BPI/CETP family domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM and SMART databases of conserved protein families/domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses, and

BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:4 is a bactericidal permeability increasing protein.

As another example, SEQ ID NO:8 is 83% identical, from residue M1 to residue M334, to human properdin (GenBank ID g35678) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 3.9e-164, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:8 also has homology to proteins that are extracellular, have structural functions, and are properdin factors, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:8 also contains a thrombospondin type 1 repeats domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based SMART database of conserved protein families/domains and a thrombospondin type 1 domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein families/domains . (See Table 3.) Data from BLIMPS and MOTIFS analyses, and BLAST analysis against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:8 is a properdin.

As another example, SEQ ID NO:23 is a splice variant of human interleukin-2 (GenBank ID g33781) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 9.3e-62, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:23 also has homology to proteins that are interleukin-2, T-cell-derived cytokines that promote activation and proliferation of lymphocytes, are involved in the immune response, are implicated in Sjorgen's syndrome, autoimmune hemolytic anemia, and multiple sclerosis, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:23 also contains an interleukin-2 domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM and SMART databases of conserved protein families/domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses, and BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:23 is an interleukin-2.

For example, SEQ ID NO:31 is a splice variant of human pentaxin (GenBank ID g35797) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.5e-131, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:31 also has homology to proteins that play roles in inflammation and the bacterial defense response, may limit autoimmune reactions during inflammation, and are members of the pentaxin family of acute-phase proteins, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:31 also contains a pentaxin family domain as determined by searching for statistically significant matches in the hidden Markov model

(HMM)-based SMART and PFAM databases of conserved protein families/domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses, and BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:31 is a pentaxin.

SEQ ID NO:1-3, SEQ ID NO:5-7, SEQ ID NO:9-22, SEQ ID NO:24-30, and SEQ ID NO:32 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-32 are described in Table 7.

As shown in Table 4, the full length polynucleotide embodiments were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide embodiments, and of fragments of the polynucleotides which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:33-64 or that distinguish between SEQ ID NO:33-64 and related polynucleotides.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotides. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as

FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotides which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotides. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

Table 8 shows single nucleotide polymorphisms (SNPs) found in polynucleotide sequences of the invention, along with allele frequencies in different human populations. Columns 1 and 2 show the polynucleotide sequence identification number (SEQ ID NO:) and the corresponding Incyte project identification number (PID) for polynucleotides of the invention. Column 3 shows the Incyte identification number for the EST in which the SNP was detected (EST ID), and column 4 shows the identification number for the SNP (SNP ID). Column 5 shows the position within the EST sequence

at which the SNP is located (EST SNP), and column 6 shows the position of the SNP within the full-length polynucleotide sequence (CB1 SNP). Column 7 shows the allele found in the EST sequence. Columns 8 and 9 show the two alleles found at the SNP site. Column 10 shows the amino acid encoded by the codon including the SNP site, based upon the allele found in the EST. Columns 11-14 show the frequency of allele 1 in four different human populations. An entry of n/d (not detected) indicates that the frequency of allele 1 in the population was too low to be detected, while n/a (not available) indicates that the allele frequency was not determined for the population.

The invention also encompasses IRAP variants. Various embodiments of IRAP variants can have at least about 80%, at least about 90%, or at least about 95% amino acid sequence identity to the IRAP amino acid sequence, and can contain at least one functional or structural characteristic of IRAP.

Various embodiments also encompass polynucleotides which encode IRAP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:33-64, which encodes IRAP. The polynucleotide sequences of SEQ ID NO:33-64, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses variants of a polynucleotide encoding IRAP. In particular, such a variant polynucleotide will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a polynucleotide encoding IRAP. A particular aspect of the invention encompasses a variant of a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NO:33-64 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:33-64. Any one of the polynucleotide variants described above can encode a polypeptide which contains at least one functional or structural characteristic of IRAP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide encoding IRAP. A splice variant may have portions which have significant sequence identity to a polynucleotide encoding IRAP, but will generally have a greater or lesser number of nucleotides due to additions or deletions of blocks of sequence arising from alternate splicing during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to a polynucleotide encoding IRAP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide encoding

IRAP. For example, a polynucleotide comprising a sequence of SEQ ID NO:34 and a polynucleotide comprising a sequence of SEQ ID NO:36 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:35 and a polynucleotide comprising a sequence of SEQ ID NO:37 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:39 and a polynucleotide comprising a sequence of SEQ ID NO:40 are splice variants of each other; and a polynucleotide comprising a sequence of SEQ ID NO:33 and a polynucleotide comprising a sequence of SEQ ID NO:47 are splice variants of each other. Any one of the splice variants described above can encode a polypeptide which contains at least one functional or structural characteristic of IRAP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding IRAP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring IRAP, and all such variations are to be considered as being specifically disclosed.

Although polynucleotides which encode IRAP and its variants are generally capable of hybridizing to polynucleotides encoding naturally occurring IRAP under appropriately selected conditions of stringency, it may be advantageous to produce polynucleotides encoding IRAP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding IRAP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of polynucleotides which encode IRAP and IRAP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic polynucleotide may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a polynucleotide encoding IRAP or any fragment thereof.

Embodiments of the invention can also include polynucleotides that are capable of hybridizing to the claimed polynucleotides, and, in particular, to those having the sequences shown in SEQ ID NO:33-64 and fragments thereof, under various conditions of stringency (Wahl, G.M. and

S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511). Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art (Ausubel et al., *supra*, ch. 7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

The nucleic acids encoding IRAP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (BD Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC

content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotides or fragments thereof which encode IRAP may be cloned in recombinant DNA molecules that direct expression of IRAP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other polynucleotides which encode substantially the same or a functionally equivalent polypeptides may be produced and used to express IRAP.

The polynucleotides of the invention can be engineered using methods generally known in the art in order to alter IRAP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of IRAP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired

properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, polynucleotides encoding IRAP may be synthesized, in whole or in part, using one or more chemical methods well known in the art (Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232). Alternatively, IRAP itself or a fragment thereof may be synthesized using chemical methods known in the art. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) Science 269:202-204). Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of IRAP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography (Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421). The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing (Creighton, *supra*, pp. 28-53).

In order to express a biologically active IRAP, the polynucleotides encoding IRAP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotides encoding IRAP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of polynucleotides encoding IRAP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where a polynucleotide sequence encoding IRAP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-

frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

Methods which are well known to those skilled in the art may be used to construct expression vectors containing polynucleotides encoding IRAP and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook and Russell, *supra*, ch. 1-4, and 8; Ausubel et al., *supra*, ch. 1, 3, and 15).

A variety of expression vector/host systems may be utilized to contain and express polynucleotides encoding IRAP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook and Russell, *supra*; Ausubel et al., *supra*; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355). Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of polynucleotides to the targeted organ, tissue, or cell population (Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5:350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:6340-6344; Buller, R.M. et al. (1985) Nature 317:813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31:219-226; Verma, I.M. and N. Somia (1997) Nature 389:239-242). The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotides encoding IRAP. For example, routine cloning, subcloning, and propagation of polynucleotides encoding IRAP can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Invitrogen). Ligation of polynucleotides encoding IRAP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of

nested deletions in the cloned sequence (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509). When large quantities of IRAP are needed, e.g. for the production of antibodies, vectors which direct high level expression of IRAP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of IRAP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign polynucleotide sequences into the host genome for stable propagation (Ausubel et al., *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184).

Plant systems may also be used for expression of IRAP. Transcription of polynucleotides encoding IRAP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection (The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196).

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, polynucleotides encoding IRAP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses IRAP in host cells (Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355).

For long term production of recombinant proteins in mammalian systems, stable expression of IRAP in cell lines is preferred. For example, polynucleotides encoding IRAP can be transformed

into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively (Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823). Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051). Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; BD Clontech), β-glucuronidase and its substrate β-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding IRAP is inserted within a marker gene sequence, transformed cells containing polynucleotides encoding IRAP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding IRAP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the polynucleotide encoding IRAP and that express IRAP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of IRAP using either

specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on IRAP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art (Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding IRAP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, polynucleotides encoding IRAP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with polynucleotides encoding IRAP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode IRAP may be designed to contain signal sequences which direct secretion of IRAP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted polynucleotides or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are

available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant polynucleotides encoding IRAP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric IRAP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of IRAP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the IRAP encoding sequence and the heterologous protein sequence, so that IRAP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In another embodiment, synthesis of radiolabeled IRAP may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

IRAP, fragments of IRAP, or variants of IRAP may be used to screen for compounds that specifically bind to IRAP. One or more test compounds may be screened for specific binding to IRAP. In various embodiments, 1, 2, 3, 4, 5, 10, 20, 50, 100, or 200 test compounds can be screened for specific binding to IRAP. Examples of test compounds can include antibodies, anticalins, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules.

In related embodiments, variants of IRAP can be used to screen for binding of test compounds, such as antibodies, to IRAP, a variant of IRAP, or a combination of IRAP and/or one or more variants IRAP. In an embodiment, a variant of IRAP can be used to screen for compounds that bind to a variant of IRAP, but not to IRAP having the exact sequence of a sequence of SEQ ID NO:1-32. IRAP variants used to perform such screening can have a range of about 50% to about

99% sequence identity to IRAP, with various embodiments having 60%, 70%, 75%, 80%, 85%, 90%, and 95% sequence identity.

In an embodiment, a compound identified in a screen for specific binding to IRAP can be closely related to the natural ligand of IRAP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified can be a natural ligand of a receptor IRAP (Howard, A.D. et al. (2001) Trends Pharmacol. Sci.22:132-140; Wise, A. et al. (2002) Drug Discovery Today 7:235-246).

In other embodiments, a compound identified in a screen for specific binding to IRAP can be closely related to the natural receptor to which IRAP binds, at least a fragment of the receptor, or a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for IRAP which is capable of propagating a signal, or a decoy receptor for IRAP which is not capable of propagating a signal (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Amgen Inc., Thousand Oaks CA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG_1 (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

In one embodiment, two or more antibodies having similar or, alternatively, different specificities can be screened for specific binding to IRAP, fragments of IRAP, or variants of IRAP. The binding specificity of the antibodies thus screened can thereby be selected to identify particular fragments or variants of IRAP. In one embodiment, an antibody can be selected such that its binding specificity allows for preferential identification of specific fragments or variants of IRAP. In another embodiment, an antibody can be selected such that its binding specificity allows for preferential diagnosis of a specific disease or condition having increased, decreased, or otherwise abnormal production of IRAP.

In an embodiment, anticalins can be screened for specific binding to IRAP, fragments of IRAP, or variants of IRAP. Anticalins are ligand-binding proteins that have been constructed based on a lipocalin scaffold (Weiss, G.A. and H.B. Lowman (2000) Chem. Biol. 7:R177-R184; Skerra, A. (2001) J. Biotechnol. 74:257-275). The protein architecture of lipocalins can include a beta-barrel having eight antiparallel beta-strands, which supports four loops at its open end. These loops form the natural ligand-binding site of the lipocalins, a site which can be re-engineered *in vitro* by amino acid substitutions to impart novel binding specificities. The amino acid substitutions can be made using methods known in the art or described herein, and can include conservative substitutions (e.g.,

substitutions that do not alter binding specificity) or substitutions that modestly, moderately, or significantly alter binding specificity.

In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit IRAP involves producing appropriate cells which express IRAP, either as a secreted protein or on the cell membrane. Preferred cells can include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing IRAP or cell membrane fractions which contain IRAP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either IRAP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with IRAP, either in solution or affixed to a solid support, and detecting the binding of IRAP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands (Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30). In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to improve or alter its ability to bind to its natural receptors (Cunningham, B.C. and J.A. Wells (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988).

IRAP, fragments of IRAP, or variants of IRAP may be used to screen for compounds that modulate the activity of IRAP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for IRAP activity, wherein IRAP is combined with at least one test compound, and the activity of IRAP in the presence of a test compound is compared with the activity of IRAP in the absence of the test compound. A change in the activity of IRAP in the presence of the test compound is indicative of a compound that modulates the activity of IRAP. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising IRAP under conditions suitable for IRAP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of IRAP

may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding IRAP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease (see, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337). For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding IRAP may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding IRAP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding IRAP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress IRAP, e.g., by secreting IRAP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of IRAP and immune response associated proteins. In addition, examples of tissues expressing IRAP can be found in Table 6 and can also be found in Example XI. Therefore, IRAP

appears to play a role in immune system, neurological, developmental, muscle, cell proliferative disorders, and disorders of lipid metabolism. In the treatment of disorders associated with increased IRAP expression or activity, it is desirable to decrease the expression or activity of IRAP. In the treatment of disorders associated with decreased IRAP expression or activity, it is desirable to increase the expression or activity of IRAP.

Therefore, in one embodiment, IRAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of IRAP. Examples of such disorders include, but are not limited to, an immune system disorder such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis,

encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, and ethanol myopathy; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia; cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and a disorder of lipid metabolism such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM_2 gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia,

hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity.

In another embodiment, a vector capable of expressing IRAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of IRAP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified IRAP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of IRAP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of IRAP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of IRAP including, but not limited to, those listed above.

In a further embodiment, an antagonist of IRAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of IRAP. Examples of such disorders include, but are not limited to, those immune system, neurological, developmental, muscle, cell proliferative disorders, and disorders of lipid metabolism described above. In one aspect, an antibody which specifically binds IRAP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express IRAP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding IRAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of IRAP including, but not limited to, those described above.

In other embodiments, any protein, agonist, antagonist, antibody, complementary sequence, or vector embodiments may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of IRAP may be produced using methods which are generally known in the art. In particular, purified IRAP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind IRAP. Antibodies to IRAP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and

fragments produced by a Fab expression library. In an embodiment, neutralizing antibodies (i.e., those which inhibit dimer formation) can be used therapeutically. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have application in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with IRAP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to IRAP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are substantially identical to a portion of the amino acid sequence of the natural protein. Short stretches of IRAP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to IRAP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce IRAP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte

population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for IRAP may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 246:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between IRAP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering IRAP epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for IRAP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of IRAP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple IRAP epitopes, represents the average affinity, or avidity, of the antibodies for IRAP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular IRAP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the IRAP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of IRAP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation

of IRAP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available (Catty, *supra*; Coligan et al., *supra*).

In another embodiment of the invention, polynucleotides encoding IRAP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding IRAP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding IRAP (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press, Totawa NJ).

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102:469-475; Scanlon, K.J. et al. (1995) FASEB J. 9:1288-1296). Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors (Miller, A.D. (1990) Blood 76:271-278; Ausubel et al., *supra*; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63:323-347). Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art (Rossi, J.J. (1995) Br. Med. Bull. 51:217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87:1308-1315; Morris, M.C. et al. (1997) Nucleic Acids Res. 25:2730-2736).

In another embodiment of the invention, polynucleotides encoding IRAP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA

93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in IRAP expression or regulation causes disease, the expression of IRAP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in IRAP are treated by constructing mammalian expression vectors encoding IRAP and introducing these vectors by mechanical means into IRAP-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J.-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of IRAP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (BD Clontech, Palo Alto CA).. IRAP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding IRAP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to IRAP expression are treated by constructing a retrovirus vector consisting of (i) the

polynucleotide encoding IRAP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. *et al.* (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In an embodiment, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding IRAP to cells which have one or more genetic abnormalities with respect to the expression of IRAP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999; Annu. Rev. Nutr. 19:511-544) and Verma, I.M. and N. Somia (1997; Nature 18:389:239-242).

In another embodiment, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding IRAP to target cells which have one or more genetic abnormalities with respect to the expression of IRAP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing IRAP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res.

169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999; J. Virol. 73:519-532) and Xu, H. et al. (1994; Dev. Biol. 163:152-161). The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another embodiment, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding IRAP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for IRAP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of IRAP-coding RNAs and the synthesis of high levels of IRAP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of IRAP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using

triplex DNA have been described in the literature (Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of RNA molecules encoding IRAP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA molecules encoding IRAP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous endonucleases.

In other embodiments of the invention, the expression of one or more selected polynucleotides of the present invention can be altered, inhibited, decreased, or silenced using RNA interference (RNAi) or post-transcriptional gene silencing (PTGS) methods known in the art. RNAi

is a post-transcriptional mode of gene silencing in which double-stranded RNA (dsRNA) introduced into a targeted cell specifically suppresses the expression of the homologous gene (i.e., the gene bearing the sequence complementary to the dsRNA). This effectively knocks out or substantially reduces the expression of the targeted gene. PTGS can also be accomplished by use of DNA or DNA fragments as well. RNAi methods are described by Fire, A. et al. (1998; Nature 391:806-811) and Gura, T. (2000; Nature 404:804-808). PTGS can also be initiated by introduction of a complementary segment of DNA into the selected tissue using gene delivery and/or viral vector delivery methods described herein or known in the art.

RNAi can be induced in mammalian cells by the use of small interfering RNA also known as siRNA. siRNA are shorter segments of dsRNA (typically about 21 to 23 nucleotides in length) that result *in vivo* from cleavage of introduced dsRNA by the action of an endogenous ribonuclease. siRNA appear to be the mediators of the RNAi effect in mammals. The most effective siRNAs appear to be 21 nucleotide dsRNAs with 2 nucleotide 3' overhangs. The use of siRNA for inducing RNAi in mammalian cells is described by Elbashir, S.M. et al. (2001; Nature 411:494-498).

siRNA can be generated indirectly by introduction of dsRNA into the targeted cell. Alternatively, siRNA can be synthesized directly and introduced into a cell by transfection methods and agents described herein or known in the art (such as liposome-mediated transfection, viral vector methods, or other polynucleotide delivery/introductory methods). Suitable siRNAs can be selected by examining a transcript of the target polynucleotide (e.g., mRNA) for nucleotide sequences downstream from the AUG start codon and recording the occurrence of each nucleotide and the 3' adjacent 19 to 23 nucleotides as potential siRNA target sites, with sequences having a 21 nucleotide length being preferred. Regions to be avoided for target siRNA sites include the 5' and 3' untranslated regions (UTRs) and regions near the start codon (within 75 bases), as these may be richer in regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP endonuclease complex. The selected target sites for siRNA can then be compared to the appropriate genome database (e.g., human, etc.) using BLAST or other sequence comparison algorithms known in the art. Target sequences with significant homology to other coding sequences can be eliminated from consideration. The selected siRNAs can be produced by chemical synthesis methods known in the art or by *in vitro* transcription using commercially available methods and kits such as the SILENCER siRNA construction kit (Ambion, Austin TX).

In alternative embodiments, long-term gene silencing and/or RNAi effects can be induced in selected tissue using expression vectors that continuously express siRNA. This can be accomplished using expression vectors that are engineered to express hairpin RNAs (shRNAs) using methods known in the art (see, e.g., Brummelkamp, T.R. et al. (2002) Science 296:550-553; and Paddison,

P.J. et al. (2002) Genes Dev. 16:948-958). In these and related embodiments, shRNAs can be delivered to target cells using expression vectors known in the art. An example of a suitable expression vector for delivery of siRNA is the PSILENCER1.0-U6 (circular) plasmid (Ambion). Once delivered to the target tissue, shRNAs are processed *in vivo* into siRNA-like molecules capable of carrying out gene-specific silencing.

In various embodiments, the expression levels of genes targeted by RNAi or PTGS methods can be determined by assays for mRNA and/or protein analysis. Expression levels of the mRNA of a targeted gene can be determined, for example, by northern analysis methods using the NORTHERNMAX-GLY kit (Ambion); by microarray methods; by PCR methods; by real time PCR methods; and by other RNA/polynucleotide assays known in the art or described herein. Expression levels of the protein encoded by the targeted gene can be determined, for example, by microarray methods; by polyacrylamide gel electrophoresis; and by Western analysis using standard techniques known in the art.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding IRAP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased IRAP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding IRAP may be therapeutically useful, and in the treatment of disorders associated with decreased IRAP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding IRAP may be therapeutically useful.

In various embodiments, one or more test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding IRAP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding IRAP are assayed

by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding IRAP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art (Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466).

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of IRAP, antibodies to IRAP, and mimetics, agonists, antagonists, or inhibitors of IRAP.

In various embodiments, the compositions described herein, such as pharmaceutical compositions, may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form.

These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery allows administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising IRAP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, IRAP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example IRAP or fragments thereof, antibodies of IRAP, and agonists, antagonists or inhibitors of IRAP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the

subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind IRAP may be used for the diagnosis of disorders characterized by expression of IRAP, or in assays to monitor patients being treated with IRAP or agonists, antagonists, or inhibitors of IRAP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for IRAP include methods which utilize the antibody and a label to detect IRAP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring IRAP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of IRAP expression. Normal or standard values for IRAP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to IRAP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of IRAP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, polynucleotides encoding IRAP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotides, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of IRAP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of

IRAP, and to monitor regulation of IRAP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotides, including genomic sequences, encoding IRAP or closely related molecules may be used to identify nucleic acid sequences which encode IRAP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding IRAP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the IRAP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:33-64 or from genomic sequences including promoters, enhancers, and introns of the IRAP gene.

Means for producing specific hybridization probes for polynucleotides encoding IRAP include the cloning of polynucleotides encoding IRAP or IRAP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides encoding IRAP may be used for the diagnosis of disorders associated with expression of IRAP. Examples of such disorders include, but are not limited to, an immune system disorder such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's

syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, and ethanol myopathy; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal

hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia; cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and a disorder of lipid metabolism such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM_2 gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity.. Polynucleotides encoding IRAP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered IRAP expression. Such qualitative or quantitative methods are well known in the art.

In a particular embodiment, polynucleotides encoding IRAP may be used in assays that detect the presence of associated disorders, particularly those mentioned above. Polynucleotides complementary to sequences encoding IRAP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of polynucleotides encoding IRAP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of IRAP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding IRAP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from

normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding IRAP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding IRAP, or a fragment of a polynucleotide complementary to the polynucleotide encoding IRAP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from polynucleotides encoding IRAP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from polynucleotides encoding IRAP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual

overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641).

Methods which may also be used to quantify the expression of IRAP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves (Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotides described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and

effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, IRAP, fragments of IRAP, or antibodies specific for IRAP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time (Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484; hereby expressly incorporated by reference herein). Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not

necessary for the statistical matching of signatures which leads to prediction of toxicity (see, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at niehs.nih.gov/oc/news/toxchip.htm). Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In an embodiment, the toxicity of a test compound can be assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another embodiment relates to the use of the polypeptides disclosed herein to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of interest. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for IRAP to quantify the levels of IRAP expression. In one embodiment, the antibodies are used as elements on a microarray,

and protein expression levels are quantified by contacting the microarray with the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/25116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well known and thoroughly described in Schena, M., ed. (1999; DNA Microarrays: A Practical

Approach, Oxford University Press, London).

In another embodiment of the invention, nucleic acid sequences encoding IRAP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; Trask, B.J. (1991) Trends Genet. 7:149-154). Once mapped, the nucleic acid sequences may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP) (Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357).

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data (Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968). Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding IRAP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation (Gatti, R.A. et al. (1988) Nature 336:577-580). The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, IRAP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a

solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between IRAP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest (Geysen, et al. (1984) PCT application WO84/03564). In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with IRAP, or fragments thereof, and washed. Bound IRAP is then detected by methods well known in the art. Purified IRAP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding IRAP specifically compete with a test compound for binding IRAP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with IRAP.

In additional embodiments, the nucleotide sequences which encode IRAP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/429,442, U.S. Ser. No. 60/429,839, U.S. Ser. No. 60/439,946 and U.S. Ser. No. 60/446,182, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs are derived from cDNA libraries described in the LIFESEQ database (Incyte, Palo Alto CA). Some tissues are homogenized and lysed in guanidinium isothiocyanate, while others are homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates are centrifuged over CsCl cushions or extracted with chloroform. RNA is precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA are repeated as necessary to increase RNA purity. In some cases, RNA is treated with DNase. For most libraries, poly(A)+ RNA is isolated

using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA is isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene is provided with RNA and constructs the corresponding cDNA libraries. Otherwise, cDNA is synthesized and cDNA libraries are constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art (Ausubel et al., *supra*, ch. 5). Reverse transcription is initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters are ligated to double stranded cDNA, and the cDNA is digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA is size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs are ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Invitrogen, Carlsbad CA), PCDNA2.1 plasmid (Invitrogen), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte, Palo Alto CA), pRARE (Incyte), or pINCY (Incyte), or derivatives thereof. Recombinant plasmids are transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Invitrogen.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I are recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids are purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids are resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA is amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps are carried out in a single reaction mixture. Samples are processed and stored in 384-well plates, and the concentration of amplified plasmid DNA is quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II are sequenced as follows.

Sequencing reactions are processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions are prepared using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides are carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences are identified using standard methods (Ausubel et al., *supra*, ch. 7). Some of the cDNA sequences are selected for extension using the techniques disclosed in Example VIII.

Polynucleotide sequences derived from Incyte cDNAs are validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof are then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz, J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families; see, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries are performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences are assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) are used to extend Incyte cDNA assemblages to full length. Assembly is performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages are screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences are translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide

sequences are subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (MiraiBio, Alameda CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences are also used to identify polynucleotide sequence fragments from SEQ ID NO:33-64. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative immune response associated proteins are initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94; Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once is set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode immune response associated proteins, the encoded polypeptides are analyzed by querying against PFAM models for immune response associated proteins. Potential immune response associated proteins are also identified by homology to Incyte cDNA sequences that have been annotated as immune response associated proteins. These selected Genscan-predicted sequences are then compared by BLAST analysis to the genpept and gbpri public databases. Where

necessary, the Genscan-predicted sequences are then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis is also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage is available, this information is used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences are obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences are derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences are extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III are mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster is analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that are subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval is present on more than one sequence in the cluster are identified, and intervals thus identified are considered to be equivalent by transitivity. For example, if an interval is present on a cDNA and two genomic sequences, then all three intervals are considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified are then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) are given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences are translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan are corrected by comparison to the top BLAST hit from genpept. Sequences are further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences are extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III are queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog is then compared by BLAST

analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein is generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both are used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences are therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences are examined to determine whether they contain a complete gene.

VI. Chromosomal Mapping of IRAP Encoding Polynucleotides

The sequences used to assemble SEQ ID NO:33-64 are compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:33-64 are assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon are used to determine if any of the clustered sequences have been previously mapped. Inclusion of a mapped sequence in a cluster results in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook and Russell, *supra*, ch. 7; Ausubel et al., *supra*, ch. 4).

Analogous computer techniques applying BLAST are used to search for identical or related molecules in databases such as GenBank or LIFESEQ (Incyte). This analysis is much faster than

multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotides encoding IRAP are analyzed with respect to the tissue sources from which they are derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding IRAP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ database (Incyte, Palo Alto CA).

VIII. Extension of IRAP Encoding Polynucleotides

Full length polynucleotides are produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer is synthesized to initiate 5' extension of the known fragment, and the other primer is synthesized to initiate 3' extension of the known fragment. The initial primers are designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations is avoided.

Selected human cDNA libraries are used to extend the sequence. If more than one extension is necessary or desired, additional or nested sets of primers are designed.

High fidelity amplification is obtained by PCR using methods well known in the art. PCR is performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contains DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ are as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well is determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate is scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture is analyzed by electrophoresis on a 1 % agarose gel to determine which reactions are successful in extending the sequence.

The extended nucleotides are desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun sequencing, the digested nucleotides are separated on low concentration (0.6 to 0.8%) agarose gels, fragments are excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected

into competent *E. coli* cells. Transformed cells are selected on antibiotic-containing media, and individual colonies are picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells are lysed, and DNA is amplified by PCR using Taq DNA polymerase (Amersham Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA is quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries are reamplified using the same conditions as described above. Samples are diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotides are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in IRAP Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) are identified in SEQ ID NO:33-64 using the LIFESEQ database (Incyte). Sequences from the same gene are clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters is used to distinguish SNPs from other sequence variants. Preliminary filters remove the majority of basecall errors by requiring a minimum Phred quality score of 15, and remove sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis is applied to the original chromatogram files in the vicinity of the putative SNP. Clone error filters use statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters use statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removes duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs are selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprises 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The African population comprises 194 individuals (97 male, 97 female), all African Americans. The

Hispanic population comprises 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprises 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies are first analyzed in the Caucasian population; in some cases those SNPs which show no allelic variance in this population are not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:33-64 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to NYTRAN PLUS nylon membranes (Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing; see, e.g., Baldeschweiler et al., *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena, M., ed. (1999) DNA Microarrays: A Practical Approach, Oxford University Press, London). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat.

Biotechnol. 16:27-31).

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (BD Clontech, Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope

slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma-Aldrich, St. Louis MO) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores.

Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte). Array elements that exhibit at least about a two-fold change in expression, a signal-to-background ratio of at least about 2.5, and an element spot size of at least about 40%, are considered to be differentially expressed.

Expression

For example, SEQ ID NO:34 showed tissue-specific expression as determined by microarray analysis. RNA samples isolated from a variety of normal human tissues were compared to a common reference sample. Tissues contributing to the reference sample were selected for their ability to provide a complete distribution of RNA in the human body and include brain (4%), heart (7%), kidney (3%), lung (8%), placenta (46%), small intestine (9%), spleen (3%), stomach (6%), testis (9%), and uterus (5%). The normal tissues assayed were obtained from at least three different donors. RNA from each donor was separately isolated and individually hybridized to the microarray.

Since these hybridization experiments were conducted using a common reference sample, differential expression values are directly comparable from one tissue to another. The expression of SEQ ID NO:34 was increased by at least two-fold in testis as compared to the reference sample. Therefore, SEQ ID NO:34 can be used as a tissue marker for testis.

SEQ ID NO:39 showed differential expression, as determined by microarray analysis, in androgen-treated (methyltrienolone (R1881), a synthetic androgen analog) human prostate tumor cells (DU-145) as compared to untreated cells. DU-145 is a prostate carcinoma cell line isolated from a metastatic site in the brain. Control and androgen-treated cells were compared at various time points. The expression of SEQ ID NO:39 was decreased by at least two-fold at the first time point only (4 hours). Therefore, in various embodiments, SEQ ID NO:39 can be used for one or more of the following: i) monitoring treatment of prostate cancer, ii) diagnostic assays for prostate cancer, and iii) developing therapeutics and/or other treatments for prostate cancer.

For example, expression of SEQ ID NO:41 was downregulated in diseased colon tissue versus normal colon tissue as determined by microarray analysis. Gene expression profiles were obtained by comparing normal colon tissue to colon adenocarcinoma tissue from the same donor (Huntsman Cancer Institute, Salt Lake City, UT) by competitive hybridization. Expression of SEQ ID NO:41 was decreased at least two-fold in colon adenocarcinoma tissue when compared to normal colon tissue from the same donor. Therefore, in various embodiments, SEQ ID NO:41 can be used for one or more of the following: i) monitoring treatment of colon cancer, ii) diagnostic assays for colon cancer, and iii) developing therapeutics and/or other treatments for colon cancer.

For example, SEQ ID NO:42 showed tissue-specific expression as determined by microarray analysis. RNA samples isolated from a variety of normal human tissues were compared to a common reference sample. Tissues contributing to the reference sample were selected for their ability to provide a complete distribution of RNA in the human body and include brain (4%), heart (7%), kidney (3%), lung (8%), placenta (46%), small intestine (9%), spleen (3%), stomach (6%), testis (9%), and uterus (5%). The normal tissues assayed were obtained from at least three different donors. RNA from each donor was separately isolated and individually hybridized to the microarray. Since these hybridization experiments were conducted using a common reference sample, differential expression values are directly comparable from one tissue to another. The expression of SEQ ID NO:42 was increased by at least two-fold in esophagus tissue as compared to the reference sample. Therefore, SEQ ID NO:42 can be used as a tissue marker for esophagus.

In another example, SEQ ID NO:42 showed tissue-specific expression as determined by microarray analysis. RNA samples isolated from a variety of normal human tissues were compared to a common reference sample. Tissues contributing to the reference sample were selected for their ability to provide a complete distribution of RNA in the human body and include brain (4%), heart

(7%), kidney (3%), lung (8%), placenta (46%), small intestine (9%), spleen (3%), stomach (6%), testis (9%), and uterus (5%). The normal tissues assayed were obtained from at least three different donors. RNA from each donor was separately isolated and individually hybridized to the microarray. Since these hybridization experiments were conducted using a common reference sample, differential expression values are directly comparable from one tissue to another. The expression of SEQ ID NO:42 was increased by at least two-fold in gallbladder tissue as compared to the reference sample. Therefore, SEQ ID NO:42 can be used as a tissue marker for gallbladder.

For example, expression of SEQ ID NO:45 was downregulated in treated breast tissue versus untreated breast tissue as determined by microarray analysis. Gene expression profiles of nonmalignant mammary epithelial cells were compared to gene expression profiles of various breast carcinoma lines at different stages of tumor progression. The cells were grown in defined serum-free H14 medium to 70-80% confluence prior to RNA harvest. Cell lines compared included: a) HMEC, a primary breast epithelial cell line isolated from a normal donor, b) MCF-10A, a breast mammary gland cell line isolated from a 36-year-old woman with fibrocystic breast disease, c) MCF7, a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69-year-old female, d) T-47D, a breast carcinoma cell line isolated from a pleural effusion obtained from a 54-year-old female with an infiltrating ductal carcinoma of the breast, e) Sk-BR-3, a breast adenocarcinoma cell line isolated from a malignant pleural effusion of a 43-year-old female, f) BT-20, a breast carcinoma cell line derived *in vitro* from cells emigrating out of thin slices of the tumor mass isolated from a 74-year-old female, g) MDA-mb-231, a breast tumor cell line isolated from the pleural effusion of a 51-year-old female, and h) MDA-mb-435S, a spindle-shaped strain that evolved from the parent line (435) isolated by R. Cailleau from pleural effusion of a 31-year-old female with metastatic, ductal adenocarcinoma of the breast.. Expression of SEQ ID NO:45 was decreased between five- and nine--fold in MCF7, T-47D, Sk-BR-3, BT-20, and MDA-mb-435S cell lines. Therefore, in various embodiments, SEQ ID NO:45 can be used for one or more of the following: i) monitoring treatment of breast cancer, ii) diagnostic assays for breast cancer, and iii) developing therapeutics and/or other treatments for breast cancer.

In yet another example, expression of SEQ ID NO:45 was upregulated in diseased colon tissue versus normal colon tissue as determined by microarray analysis. Gene expression profiles were obtained by comparing normal sigmoid colon tissue from a donor to a sigmoid colon tumor originating from a metastatic gastric sarcoma (stromal tumor) from the same donor (Huntsman Cancer Institute, Salt Lake City, UT). Expression of SEQ ID NO:45 was increased at least two-fold in the colon sigmoid colon tumor when compared to normal sigmoid colon tissue from the same donor. Therefore, in various embodiments, SEQ ID NO:45 can be used for one or more of the following: i) monitoring treatment of colon cancer, ii) diagnostic assays for colon cancer, and iii)

developing therapeutics and/or other treatments for colon cancer.

In a further example, expression of SEQ ID NO:45 was upregulated in diseased ovarian tissue versus normal ovarian tissue as determined by microarray analysis. A normal ovary was compared to an ovarian tumor from the same donor (Huntsman Cancer Institute, Salt Lake City, UT). Expression of SEQ ID NO:45 was increased at least two-fold inthe ovarian tumor tissue when compared to normal ovarian tissue from the same donor. Therefore, in various embodiments, SEQ ID NO:45 can be used for one or more of the following: i) monitoring treatment of ovarian cancer, ii) diagnostic assays for ovarian cancer, and iii) developing therapeutics and/or other treatments for ovarian cancer.

For example, expression of SEQ ID NO:53 was down-regulated in colon tumor tissue versus normal colon tissue as determined by microarray analysis. Expression of SEQ ID NO:53 was decreased at least two-fold in matched colon tumor tissue versus normal colon tissue from 5 of 14 donors. Therefore, in various embodiments, SEQ ID NO:53 can be used for one or more of the following: i) monitoring treatment of colon cancer, ii) diagnostic assays for colon cancer, and iii) developing therapeutics and/or other treatments for colon cancer.

In another example, expression of SEQ ID NO:53 was up-regulated in lung tumor tissue versus normal lung tissue as determined by microarray analysis. Expression of SEQ ID NO:53 was increased at least two-fold in matched lung tumor tissue versus normal lung tissue from 1 of 4 donors. Therefore, in various embodiments, SEQ ID NO:53 can be used for one or more of the following: i) monitoring treatment of lung cancer, ii) diagnostic assays for lung cancer, and iii) developing therapeutics and/or other treatments for lung cancer.

In another example, expression of SEQ ID NO:56 was down-regulated in breast carcinoma cell lines versus cells derived from non-malignant fibrocystic breast epithelial cells as determined by microarray analysis. Gene expression profiles of nonmalignant mammary epithelial cells were compared to gene expression profiles of various breast carcinoma lines at different stages of tumor progression. The cells were grown in defined serum-free H14 medium to 70-80% confluence prior to RNA harvest. Cell lines compared included: a) MCF-10A, a breast mammary gland (luminal ductal characteristics) cell line isolated from a 36-year-old woman with fibrocystic breast disease; b) MCF7, a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69-year-old female, c) T-47D, a breast carcinoma cell line isolated from a pleural effusion obtained from a 54-year-old female with an infiltrating ductal carcinoma of the breast, d) Sk-BR-3, a breast adenocarcinoma cell line isolated from a malignant pleural effusion of a 43-year-old female, e) BT-20, a breast carcinoma cell line derived *in vitro* from the cells emigrating out of thin slices of the tumor mass isolated from a 74-year-old female, and f) MDA-mb-231, a breast tumor cell line isolated from the pleural effusion of a 51-year old female. Expression of SEQ ID NO:56 was

decreased at least two-fold in 3 breast carcinoma cell lines (T-47D, BT-20, and MCF7) versus the non-malignant cells (MCF-10A). Therefore, in various embodiments, SEQ ID NO:56 can be used for one or more of the following: i) monitoring treatment of breast cancer, ii) diagnostic assays for breast cancer, and iii) developing therapeutics and/or other treatments for breast cancer.

In another example, expression of SEQ ID NO:56 was down-regulated in breast tumor tissue versus normal breast tissue as determined by microarray analysis. Expression of SEQ ID NO:56 was decreased at least two-fold in matched breast lobular carcinoma tumor tissue versus normal breast tissue from one donor. Therefore, in various embodiments, SEQ ID NO:56 can be used for one or more of the following: i) monitoring treatment of breast cancer, ii) diagnostic assays for breast cancer, and iii) developing therapeutics and/or other treatments for breast cancer.

In another example, expression of SEQ ID NO:56 was down-regulated in ovarian tumor tissue versus normal ovarian tissue as determined by microarray analysis. Expression of SEQ ID NO:56 was decreased at least two-fold in matched ovarian tumor tissue versus normal ovarian tissue from one donor. Therefore, in various embodiments, SEQ ID NO:56 can be used for one or more of the following: i) monitoring treatment of ovarian cancer, ii) diagnostic assays for ovarian cancer, and iii) developing therapeutics and/or other treatments for ovarian cancer.

For example, expression of SEQ ID NO:56 was up-regulated in a prostate cancer cell line versus cells dervived from normal prostate epithelium as determined by microarray analysis. Primary prostate epithelial cells were compared with prostate carcinomas representative of the different stages of tumor progression. Cell lines compared included: a) PrEC, a primary prostate epithelial cell line isolated from a normal donor, b) DU 145, a prostate carcinoma cell line isolated from a metastatic site in the brain of 69-year old male with widespread metastatic prostate carcinoma, c) LNCaP, a prostate carcinoma cell line isolated from a lymph node biopsy of a 50-year-old male with metastatic prostate carcinoma, and d) PC-3, a prostate adenocarcinoma cell line isolated from a metastatic site in the bone of a 62-year-old male with grade IV prostate adenocarcinoma. Cell lines were grown in basal media in the absence of growth factors and hormones and were compared to normal PrECs grown under the same conditions. Expression of SEQ ID NO:56 was increased at least two-fold in one prostate cancer cell line (DU 145) versus PrECs. Therefore, in various embodiments, SEQ ID NO:56 can be used for one or more of the following: i) monitoring treatment of prostate cancer, ii) diagnostic assays for prostate cancer and iii) developing therapeutics and/or other treatments for prostate cancer.

In two other examples, SEQ ID NO:53 and SEQ ID NO:58 showed tissue-specific expression as determined by microarray analysis. RNA samples isolated from a variety of normal human tissues were compared to a common reference sample. Tissues contributing to the reference sample were selected for their ability to provide a complete distribution of RNA in the human body

and include brain (4%), heart (7%), kidney (3%), lung (8%), placenta (46%), small intestine (9%), spleen (3%), stomach (6%), testis (9%), and uterus (5%). The normal tissues assayed were obtained from at least three different donors. RNA from each donor was separately isolated and individually hybridized to the microarray. Since these hybridization experiments were conducted using a common reference sample, differential expression values are directly comparable from one tissue to another. In one example the expression of SEQ ID NO:53 was increased by at least two-fold in salivary gland as compared to the reference sample. Therefore, SEQ ID NO:53 can be used as a tissue marker for salivary gland. In a second example, SEQ ID NO:58 was increased by at least two-fold in spleen and tonsils as compared to the reference sample. Therefore, SEQ ID NO:58 can be used as a tissue marker for spleen and tonsils.

For example, expression of SEQ ID NO:63 showed differential expression in breast carcinoma cell lines compared with a primary culture of epithelial cells derived from normal breast tissue as determined by microarray analysis. The gene expression profile of a nonmalignant mammary epithelial cell line was compared to the gene expression profiles of breast carcinoma lines at different stages of tumor progression. Cell lines compared included: a) BT-20, a breast carcinoma cell line derived *in vitro* from the cells emigrating out of thin slices of tumor mass isolated from a 74-year-old female, b) BT-474, a breast ductal carcinoma cell line that was isolated from a solid, invasive ductal carcinoma of the breast obtained from a 60-year-old woman, c) BT-483, a breast ductal carcinoma cell line that was isolated from a papillary invasive ductal tumor obtained from a 23-year-old normal, menstruating, parous female with a family history of breast cancer, d) Hs 578T, a breast ductal carcinoma cell line isolated from a 74-year-old female with breast carcinoma, e) MCF7, a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69-year-old female, f) MCF-10A, a breast mammary gland (luminal ductal characteristics) cell line isolated from a 36-year-old woman with fibrocystic breast disease, g) MDA-MB-468, a breast adenocarcinoma cell line isolated from the pleural effusion of a 51-year-old female with metastatic adenocarcinoma of the breast, and h) HMEC, a primary breast epithelial cell line isolated from a normal donor. Expression of SEQ ID NO:63 was increased at least two-fold in one breast cancer cell line (Hs 578T) and decreased at least two-fold in one breast cancer cell line (BT-474. Although expression of SEQ ID NO:63 was not affected in the same manner among all breast carcinoma cell lines, the data suggest that in some populations or stages of breast cancer SEQ ID NO:63 is differentially expressed. Therefore, in various embodiments, SEQ ID NO:63 can be used for one or more of the following: i) monitoring treatment of breast cancer, ii) diagnostic assays for breast cancer, and iii) developing therapeutics and/or other treatments for breast cancer.

For example, expression of SEQ ID NO:64 was down-regulated in preadipocytes taken from an obese donor versus preadipocytes taken from a non-obese donor as determined by microarray

analysis. Primary subcutaneous preadipocytes were isolated from the adipose tissue of a non-obese donor, a 28-year-old healthy female with body mass index (BMI) of 23.59, and an obese donor, a 40-year-old healthy female with a body mass index (BMI) of 32.47. The preadipocytes from each donor were cultured and induced to differentiate into adipocytes by growing them in differentiation medium containing PPAR-γ agonist and human insulin (Zen-Bio). Some thiazolidinediones or PPAR-γ agonists, which bind and activate an orphan nuclear receptor, PPAR-γ, have been shown to induce human adipocyte differentiation. The preadipocytes were treated with human insulin and PPAR-γ agonist for 3 days and subsequently were switched to medium containing insulin for a range of time periods ranging from one to 20 days before the cells were collected for analysis. Differentiated adipocytes from each donor were compared to untreated preadipocytes, maintained in culture in the absence of differentiation-inducing agents, from the same donor. Between 80% and 90% of the preadipocytes finally differentiated to adipocytes as observed under phase contrast microscopy. Expression of SEQ ID NO:64 was decreased at least two-fold in differentiated preadipocytes from an obese donor versus non-differentiated preadipocytes from the same donor. In contrast, no differential expression was seen in differentiated preadipocytes from a non-obese donor versus non-differentiated preadipocytes from the same donor. These data suggest that SEQ ID NO:64 is differentially expressed in adipocytes from obese subjects but not in adipocytes from non-obese subjects. Thus, SEQ ID NO:64 is useful for the diagnosis, prognosis, or treatment of diabetes mellitus, obesity, hypertension, and atherosclerosis. Therefore, in various embodiments, SEQ ID NO:64 can be used for one or more of the following: i) monitoring treatment of diabetes mellitus, obesity, hypertension, and atherosclerosis, ii) diagnostic assays for diabetes mellitus, obesity, hypertension, and atherosclerosis, and iii) developing therapeutics and/or other treatments for diabetes mellitus, obesity, hypertension, and atherosclerosis.

In another example, SEQ ID NO:61 showed tissue-specific expression as determined by microarray analysis. RNA samples isolated from a variety of normal human tissues were compared to a common reference sample. Tissues contributing to the reference sample were selected for their ability to provide a complete distribution of RNA in the human body and include brain (4%), heart (7%), kidney (3%), lung (8%), placenta (46%), small intestine (9%), spleen (3%), stomach (6%), testis (9%), and uterus (5%). The normal tissues assayed were obtained from at least three different donors. RNA from each donor was separately isolated and individually hybridized to the microarray. Since these hybridization experiments were conducted using a common reference sample, differential expression values are directly comparable from one tissue to another. The expression of SEQ ID NO:61 was increased by at least two-fold in thymus gland as compared to the reference sample. Therefore, SEQ ID NO:61 can be used as a tissue marker for thymus gland.

XII. Complementary Polynucleotides

Sequences complementary to the IRAP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring IRAP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of IRAP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the IRAP-encoding transcript.

XIII. Expression of IRAP

Expression and purification of IRAP is achieved using bacterial or virus-based expression systems. For expression of IRAP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express IRAP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of IRAP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding IRAP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus (Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945).

In most expression systems, IRAP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences). Following purification, the GST moiety can be proteolytically cleaved from IRAP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification

using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). Purified IRAP obtained by these methods can be used directly in the assays shown in Examples XVII and XVIII, where applicable.

XIV. Functional Assays

IRAP function is assessed by expressing the sequences encoding IRAP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; BD Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994; Flow Cytometry, Oxford, New York NY).

The influence of IRAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding IRAP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding IRAP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of IRAP Specific Antibodies

IRAP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the IRAP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art (Ausubel et al., *supra*, ch. 11).

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity (Ausubel et al., *supra*). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-IRAP activity by, for example, binding the peptide or IRAP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring IRAP Using Specific Antibodies

Naturally occurring or recombinant IRAP is substantially purified by immunoaffinity chromatography using antibodies specific for IRAP. An immunoaffinity column is constructed by covalently coupling anti-IRAP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing IRAP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of IRAP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/IRAP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and IRAP is collected.

XVII. Identification of Molecules Which Interact with IRAP

IRAP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent (Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled IRAP, washed, and any wells with labeled IRAP complex are assayed. Data obtained using different concentrations of IRAP are used to calculate values for the number, affinity, and association of IRAP with the candidate molecules.

Alternatively, molecules interacting with IRAP are analyzed using the yeast two-hybrid

system as described in Fields, S. and O. Song (1989; Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (BD Clontech).

IRAP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of IRAP Activity

An assay for IRAP activity measures the ability of IRAP to recognize and precipitate antigens from serum. This activity can be measured by the quantitative precipitin reaction (Golub, E.S. et al. (1987) Immunology: A Synthesis, Sinauer Associates, Sunderland, MA, pages 113-115). IRAP is isotopically labeled using methods known in the art. Various serum concentrations are added to constant amounts of labeled IRAP. IRAP-antigen complexes precipitate out of solution and are collected by centrifugation. The amount of precipitable IRAP-antigen complex is proportional to the amount of radioisotope detected in the precipitate. The amount of precipitable IRAP-antigen complex is plotted against the serum concentration. For various serum concentrations, a characteristic precipitin curve is obtained, in which the amount of precipitable IRAP-antigen complex initially increases proportionately with increasing serum concentration, peaks at the equivalence point, and then decreases proportionately with further increases in serum concentration. Thus, the amount of precipitable IRAP-antigen complex is a measure of IRAP activity which is characterized by sensitivity to both limiting and excess quantities of antigen.

Alternatively, an assay for IRAP activity measures the expression of IRAP on the cell surface. cDNA encoding IRAP is transfected into a non-leukocytic cell line. Cell surface proteins are labeled with biotin (de la Fuente, M.A. et al. (1997) Blood 90:2398-2405). Immunoprecipitations are performed using IRAP-specific antibodies, and immunoprecipitated samples are analyzed using SDS-PAGE and immunoblotting techniques. The ratio of labeled immunoprecipitant to unlabeled immunoprecipitant is proportional to the amount of IRAP expressed on the cell surface.

Alternatively, an assay for IRAP activity measures the amount of cell aggregation induced by overexpression of IRAP. In this assay, cultured cells such as NIH3T3 are transfected with cDNA encoding IRAP contained within a suitable mammalian expression vector under control of a strong promoter. Cotransfection with cDNA encoding a fluorescent marker protein, such as Green Fluorescent Protein (CLONTECH), is useful for identifying stable transfectants. The amount of cell agglutination, or clumping, associated with transfected cells is compared with that associated with untransfected cells. The amount of cell agglutination is a direct measure of IRAP activity.

IRAP protease activity is measured by the hydrolysis of appropriate synthetic peptide

substrates conjugated with various chromogenic molecules in which the degree of hydrolysis is quantified by spectrophotometric (or fluorometric) absorption of the released chromophore (Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York, NY, pp. 25-55). Peptide substrates are designed according to the category of protease activity as endopeptidase (serine, cysteine, aspartic proteases, or metalloproteases), aminopeptidase (leucine aminopeptidase), or carboxypeptidase (carboxypeptidases A and B, procollagen C-proteinase). Commonly used chromogens are 2-naphthylamine, 4-nitroaniline, and furylacrylic acid. Assays are performed at ambient temperature and contain an aliquot of the enzyme and the appropriate substrate in a suitable buffer. Reactions are carried out in an optical cuvette, and the increase/decrease in absorbance of the chromogen released during hydrolysis of the peptide substrate is measured. The change in absorbance is proportional to the enzyme activity in the assay.

In the alternative, an assay for IRAP protease activity takes advantage of fluorescence resonance energy transfer (FRET) that occurs when one donor and one acceptor fluorophore with an appropriate spectral overlap are in close proximity. A flexible peptide linker containing a cleavage site specific for IRAP is fused between a red-shifted variant (RSGFP4) and a blue variant (BFP5) of Green Fluorescent Protein. This fusion protein has spectral properties that suggest energy transfer is occurring from BFP5 to RSGFP4. When the fusion protein is incubated with IRAP, the substrate is cleaved, and the two fluorescent proteins dissociate. This is accompanied by a marked decrease in energy transfer which is quantified by comparing the emission spectra before and after the addition of IRAP (Mitra, R.D. et al (1996) Gene 173:13-17). This assay can also be performed in living cells. In this case the fluorescent substrate protein is expressed constitutively in cells and IRAP is introduced on an inducible vector so that FRET can be monitored in the presence and absence of IRAP (Sagot, I. et al (1999) FEBS Letters 447:53-57).

Various modifications and variations of the described compositions, methods, and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. It will be appreciated that the invention provides novel and useful proteins, and their encoding polynucleotides, which can be used in the drug discovery process, as well as methods for using these compositions for the detection, diagnosis, and treatment of diseases and conditions. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Nor should the description of such embodiments be considered exhaustive or limit the invention to the precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined with elements from one or more other embodiments. Such combinations can form a number of embodiments within the scope of the invention. It is intended that the scope of the invention be defined by the following claims and their equivalents.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
7522043	1	7522043CD1	33	7522043CB1
7523539	2	7523539CD1	34	7523539CB1
7523587	3	7523587CD1	35	7523587CB1
7523622	4	7523622CD1	36	7523622CB1
7523711	5	7523711CD1	37	7523711CB1
7523729	6	7523729CD1	38	7523729CB1
7523763	7	7523763CD1	39	7523763CB1
7523006	8	7523006CD1	40	7523006CB1
7523261	9	7523261CD1	41	7523261CB1
7523277	10	7523277CD1	42	7523277CB1
7523279	11	7523279CD1	43	7523279CB1
7523296	12	7523296CD1	44	7523296CB1
7521779	13	7521779CD1	45	7521779CB1
7521826	14	7521826CD1	46	7521826CB1
7521901	15	7521901CD1	47	7521901CB1
7522003	16	7522003CD1	48	7522003CB1
7522014	17	7522014CD1	49	7522014CB1
7522038	18	7522038CD1	50	7522038CB1
7523429	19	7523429CD1	51	7523429CB1
7523941	20	7523941CD1	52	7523941CB1
7524607	21	7524607CD1	53	7524607CB1
7524690	22	7524690CD1	54	7524690CB1
7524733	23	7524733CD1	55	7524733CB1
7522128	24	7522128CD1	56	7522128CB1
7522158	25	7522158CD1	57	7522158CB1
7524191	26	7524191CD1	58	7524191CB1
7525225	27	7525225CD1	59	7525225CB1
7513053	28	7513053CD1	60	7513053CB1
7513086	29	7513086CD1	61	7513086CB1
7513557	30	7513557CD1	62	7513557CB1
7513718	31	7513718CD1	63	7513718CB1
7514003	32	7514003CD1	64	7514003CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	7522043CD1	g13021810	1.7E-115	[Homo sapiens] NK cell receptor
				Boles, K. S. et al., Molecular cloning of CS1, a novel human natural killer cell receptor belonging to the CD2 subset of the immunoglobulin superfamily. Immunogenetics 52, 302-307 (2001)
	7522043CD1	789879|CRACC	1.3E-116	[Homo sapiens] Protein with low similarity to cluster of differentiation antigen 84 (human CD84), which binds the X-linked lymphoproliferative disease gene product SH2D1A (human SAP) and functions in the immune response, cell adhesion, and possibly leukocyte activation
	7522043CD1	749150|CD84-H1	9.1E-27	[Homo sapiens] Protein containing two immunoglobulin (Ig) domains, which may be involved in protein-protein and protein-ligand interactions, has low similarity to cluster of differentiation antigen 84 (human CD84), which is involved in the immune response
	7522043CD1	347606|LY9	9.5E-26	[Homo sapiens][Adhesin/agglutinin][Plasma membrane] Lymphocyte antigen 9, a member of the immunoglobulin superfamily, may be involved in adhesion reactions between T cells and accessory cells; may play a role in X linked lymphoproliferative disease
				Kingsmore, S. F. et al. Physical and genetic linkage of the genes encoding Ly-9 and CD48 on mouse and human chromosomes 1. Immunogenetics 42, 59-62 (1995)
	7522043CD1	568038|CD84	6.4E-26	[Homo sapiens][Plasma membrane] Cluster of differentiation antigen 84, member of the CD2 subset of immunoglobulin cell surface receptors, involved in immune response and cell adhesion, may play a role in leukocyte activation, interacts with the XLP disease gene product SH2D1A (SAP)
				Tangye, S. G. et al. 2B4-mediated activation of human natural killer cells. Mo. Immunol. 37, 493-501 (2000)
2	7523539CD1	g179529	2.9E-161	[Homo sapiens] bactericidal permeability increasing protein (BPI) precursor
				Gray, P. W. et al. Cloning of the cDNA of a human neutrophil bactericidal protein. Structural and functional correlations. J. Biol. Chem. 264, 9505-9509 (1989)
3	7523587CD1	g10719608	2.2E-60	[Homo sapiens] IL-22 receptor
	7523587CD1	618086|IL22R	1.6E-61	[Homo sapiens][Receptor (signalling)][Plasma membrane;Unspecified membrane] Interleukin 22 receptor, a member of the interferon receptor family, acts with IL-10R2 (IL10RB) to mediate IL22 signaling and activation of STAT protein

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Xie, M. H. et al. Interleukin (IL)-22, a novel human cytokine that signals through the interferon receptor-related proteins CRF2-4 and IL-22R. J. Biol. Chem. 275, 31335-31339 (2000)
	7523587CD1	752461|IL22RA2	2.7E-15	[Homo sapiens] Interleukin 22 (IL22) receptor-alpha 2 (IL22 binding protein), an interleukin 22 binding protein that acts as an IL-22 antagonist and may play a role in regulating inflammatory responses
				Xu, W. et al. A soluble class II cytokine receptor, IL-22RA2, is a naturally occurring IL-22 antagonist. Proc. Natl. Acad. Sci. U. S. A. 98, 9511-9516 (2001)
4	7523622CD1	g179529	1.7E-236	[Homo sapiens] bactericidal permeability increasing protein (BPI) precursor
				Gray, P. W. et al. Cloning of the cDNA of a human neutrophil bactericidal protein. Structural and functional correlations. J. Biol. Chem. 264, 9505-9509 (1989)
	7523622CD1	339686|BPI	1.3E-237	[Homo sapiens][Plasma membrane] Bactericidal permeability-increasing protein, binds to gram-negative bacteria and blocks LPS-mediated cytokine release, mediates opsinization of bacteria, can inhibit angiogenesis, and plays a role in acute pancreatitis, enteritis and cirrhosis
				Bouma, M. G. et al. Adenosine inhibits neutrophil degranulation in activated human whole blood: involvement of adenosine A2 and A3 receptors. J. Immunol. 158, 5400-5408 (1997)
	7523622CD1	724558|1ewf_A	2.4E-221	[Protein Data Bank] Bactericidal/Permeability-Increasing Protein
	7523622CD1	719814|1bp1_	6.4E-221	[Protein Data Bank] Bactericidal/Permeability-Increasing Protein
	7523622CD1	625791|LBP	2.4E-104	[Homo sapiens][Dense bodies;Extracellular (excluding cell wall)] Lipopolysaccharide (LPS)-binding protein, an acute phase protein with bactericidal activity against gram-negative bacteria, binds to the lipid A moiety of bacterial LPS, mediates proinflammatory effects of LPS by enhancing LPS-induced cytokine production
				Hubacek, J. A. et al. The genomic organization of the genes for human lipopolysaccharide binding protein (LBP) and bactericidal permeability increasing protein (BPI) is highly conserved. Biochem. Biophys. Res. Commun. 236, 427-430 (1997)
				Schumann, R. R. et al. Structure and function of lipopolysaccharide binding protein. Science 249, 1429-1431 (1990)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
	7523622CD1	591193|Lbp	6.4E-102	[Rattus norvegicus] Lipopolysaccharide binding protein, a serum glycoprotein expressed during the acute phase response, may strengthen the host response to endotoxin or lipopolysaccharide
				Su, G. L. et al. Molecular cloning, characterization, and tissue distribution of rat lipopolysaccharide binding protein. Evidence for extrahepatic expression. J. Immunol. 153, 7437-52 (1994)
5	7523711CD1	g10719608	2.5E-26	[Homo sapiens] IL-22 receptor
				Xie, M. H. et al. Interleukin (IL)-22, a novel human cytokine that signals through the interferon receptor-related proteins CRF2-4 and IL-22R. J. Biol. Chem. 275, 31335-31339 (2000)
	7523711CD1	618086|IL22R	1.8E-27	[Homo sapiens][Receptor (signalling)][Plasma membrane;Unspecified membrane] Interleukin 22 receptor, a member of the interferon receptor family, acts with IL-10R2 (IL10RB) to mediate IL22 signaling and activation of STAT protein
				Kotenko, S. V. et al. Identification of the functional interleukin-22 (IL-22) receptor complex: the IL-10R2 chain (IL-10Rbeta) is a common chain of both the IL-10 and IL-22 (IL-10-related T cell-derived inducible factor, IL-TIF) receptor complexes. J. Biol. Chem. 276, 2725-2732 (2001)
6	7523729CD1	g3153241	5.4E-38	[Homo sapiens] class I cytokine receptor
				Sprecher, C. A. et al. Cloning and characterization of a novel class I cytokine receptor. Biochem. Biophys. Res. Commun. 246, 82-90 (1998)
	7523729CD1	341448|WSX1	3.8E-39	[Homo sapiens][Receptor (signalling)][Plasma membrane] Cytokine receptor family class 1 member 1, a putative transmembrane receptor that may be involved in regulation of the immune response; mouse Wsx1 is involved in the response to Leishmania major and mycobacterial infection
				Chen, Q. et al. Development of Th1-type immune responses requires the type I cytokine receptor TCCR. Nature 407, 916-920 (2000)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
	7523729CD1	587585|Tccr	1.8E-15	[Mus musculus][Receptor (signalling)][Plasma membrane] Cytokine receptor family class 1 member 1, a protein involved in Th1 cell differentiation and IFN-gamma (Ifng) production, may also play a role in inhibition of T cell proliferation; involved in the response to Leishmania major and mycobacterial infection
				Yoshida, H. et al. WSX-1 is required for the initiation of Th1 responses and resistance to L. major infection. Immunity 15, 569-578 (2001)
7	7523763CD1	g35678	4.4E-39	[Homo sapiens] properdin
				Nolan, K. F. et al. Molecular cloning of the cDNA coding for properdin, a positive regulator of the alternative pathway of human complement. Eur. J. Immunol. 21, 771-776
				Nolan, K. F. et al., Characterization of the human properdin gene, Biochem. J. 287 (Pt 1), 291-297 (1992)
	7523763CD1	344704|PFC	3.2E-40	[Homo sapiens][Structural protein][Extracellular (excluding cell wall)] Properdin P factor, a serum protein with a related type-I repeat sequence (TSR), plays a role complement-mediated clearance and inactivation mechanisms of natural and acquired resistance to infection; deficiency leads to fatal bacterial infections
				Goundis, D. et al. Properdin, the terminal complement components, thrombospondin and the circumsporozoite protein of malaria parasites contain similar sequence motifs. Nature 335, 8285 (1988)
	7523763CD1	323766|Pfc	4.1E-16	[Mus musculus][Structural protein] Properdin factor (complement), may play a role in complement-mediated clearance through the alternative complement pathway; deficiency of human PFC leads to fatal bacterial infections
				Goundis, D. et al. Properdin, the terminal complement components, thrombospondin and the circumsporozoite protein of malaria parasites contain similar sequence motifs. Nature 335, 82-85 (1988)
8	7523006CD1	g35678	3.9E-164	[Homo sapiens] properdin
				Nolan, K. F. et al., Molecular cloning of the cDNA coding for properdin, a positive regulator of the alternative pathway of human complement, Eur. J. Immunol. 21, 771-776 (1991)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Nolan, K. F. et al., Characterization of the human properdin gene, Biochem. J. 287 (Pt 1), 291-297 (1992)
		344704|PFC	2.9E-165	[Homo sapiens][Structural protein][Extracellular (excluding cell wall)] Properdin P factor, a serum protein with a related type-I repeat sequence (TSR), plays a role in complement-mediated clearance and inactivation mechanisms of natural and acquired resistance to infection; deficiency leads to fatal bacterial infections
				Schwaeble, W. et al., Properdin, a positive regulator of complement activation, is expressed in human T cell lines and peripheral blood T cells., J Immunol 151, 2521-8. (1993).
		323766|Pfc	1.6E-118	[Mus musculus][Structural protein] Properdin factor (complement), may play a role in complement-mediated clearance through the alternative complement pathway; deficiency of human PFC leads to fatal bacterial infections
				Goundis, D. et al., Properdin, the terminal complement components, thrombospondin and the circumsporozoite protein of malaria parasites contain similar sequence motifs., Nature 335, 82-5 (1988).
9	7523261CD1	g2842401	9.7E-90	[Homo sapiens] lymphocyte function associated antigen-3, GPI-linked precursor
				Wallich, R. et al., Gene structure, promoter characterization, and basis for alternative mRNA splicing of the human CD58 gene, J. Immunol. 160, 2862-2871 (1998)
		334544|CD58	7.1E-91	[Homo sapiens][Adhesin/agglutinin][Plasma membrane] Lymphocyte function-associated antigen 3 (CD58 antigen), member of the immunoglobulin superfamily, acts as a cell surface ligand for the CD2 receptor, mediates adhesion of helper T cells to antigen presenting cells, and cytolytic effectors to target cells
				Wang, J. H. et al., Structure of a heterophilic adhesion complex between the human CD2 and CD58 (LFA-3) counterreceptors., Cell 97, 791-803 (1999).
10	7523277CD1	g2909826	6.4E-22	[Homo sapiens] erythrocyte membrane glycoprotein Rh50
				Huang, C. H., The human Rh50 glycoprotein gene. Structural organization and associated splicing defect resulting in Rh(null) disease, J. Biol. Chem. 273, 2207-2213 (1998)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		337618|RHAG	7.8E-23	[Homo sapiens][Plasma membrane] Rhesus blood group-associated glycoprotein, a component of the Rh antigen, plays a role in the antigen transport to the cell surface, and may play a role in ammonium transport; mutation in the corresponding gene causes Rh deficiency and Rh-mod syndrome
				Cherif-Zahar, B. et al., Rh-deficiency of the regulator type caused by splicing mutations in the human RH50 gene., Blood 92, 2535-40. (1998).
11	7523279CD1	g291929	1.9E-77	[Homo sapiens] cytotoxic T-lymphocyte-associated protein 4
				Dariavach, P. et al., Human Ig superfamily CTLA-4 gene: chromosomal localization and identity of protein sequence between murine and human CTLA-4 cytoplasmic domains, Eur. J. Immunol. 18, 1901-1905 (1988)
				Harper, K. et al., CTLA-4 and CD28 activated lymphocyte molecules are closely related in both mouse and human as to sequence, message expression, gene structure, and chromosomal location, J. Immunol. 147, 1037-1044 (1991)
		341996|CTLA4	1.4E-78	[Homo sapiens][Plasma membrane] Cytotoxic T-lymphocyte-associated protein 4, member of the immunoglobulin superfamily that negatively regulates T cell activation, involved in immune tolerance and plays a role in autoimmune disease, graft reactions, and tumor immunity
				Kouki, T. et al., CTLA-4 gene polymorphism at position 49 in exon 1 reduces the inhibitory function of CTLA-4 and contributes to the pathogenesis of Graves' disease, J Immunol 165, 6606-11 (2000).
		711598|Cd152	1.0E-52	[Rattus norvegicus][Plasma membrane] CD152 antigen (cytotoxic T-lymphocyte-associated protein 4), human CTLA4 plays a role in autoimmune disease, graft reactions, and tumor immunity
				Brunet, J. F. et al., A new member of the immunoglobulin superfamily--CTLA-4., Nature 328, 267-70 (1987).
12	7523296CD1	g7959947	5.8E-28	[Equus caballus] low-affinity IgE receptor; CD23
				Watson, J. L. et al., Molecular cloning and sequencing of the low-affinity IgE receptor (CD23) for horse and cattle, Vet. Immunol. Immunopathol. 73, 323-329 (2000)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		586979|Fcer2a	2.5E-30	[Mus musculus][Receptor (signaling)][Plasma membrane] Fc receptor IgE low affinity II alpha polypeptide, low affinity IgE receptor, plays a major role in trapping and retaining IgE immune complexes, involved in IgE homeostasis and antigen presentation; contributes to pathology of collagen-induced arthritis
				Gollnick, S. O. et al., Isolation, characterization, and expression of cDNA clones encoding the mouse Fc receptor for IgE (Fc epsilon RII)1., J Immunol 144, 1974-82 (1990).
		800185|FCER2	8.2E-26	[Homo sapiens][Ligand; Receptor (signaling)][Plasma membrane] Low affinity II receptor for Fc epsilon RII, contains an inverse RGD motif that binds IgE and regulates IgE synthesis, stimulates B cell proliferation and aggregation, participates in inflammatory responses associated with allergies, including asthma
				Bettler, B. et al., Molecular structure and expression of the murine lymphocyte low- affinity receptor for IgE (Fc epsilon RII)., Proc Natl Acad Sci U S A 86, 7566-70 (1989).
13	7521779CD1	g1877212	4.1E-162	[Homo sapiens] IL-13 receptor
				Guo, J. et al., Chromosome mapping and expression of the human interleukin-13 receptor, Genomics 42, 141-145 (1997)
		618416|IL13RA2	3.0E-163	[Homo sapiens][Extracellular (excluding cell wall)] Interleukin 13 receptor alpha 2, a high affinity interleukin-13 (IL13) receptor that is also involved in the internalization of IL13, may play a role in the inflammatory response
				Doucet, C. et al., Interleukin (IL) 4 and IL-13 act on human lung fibroblasts. Implication in asthma., J Clin Invest 101, 2129-39. (1998).
		583357|Il13ra2	3.2E-102	[Mus musculus][Extracellular (excluding cell wall)] Interleukin 13 receptor alpha 2, a high affinity interleukin-13 (Il13) receptor, may be involved in the inflammatory and immune responses; overexpressed in several head and neck cancer cell lines and may be a target for head and neck tumor therapy
				Kawakami, K. et al., Interleukin-13 receptor-targeted cancer therapy in an immunodeficient animal model of human head and neck cancer., Cancer Res 61, 6194-200. (2001).
14	7521826CD1	g1401185	3.0E-111	[Homo sapiens] DNAM-1
				Shibuya, A. et al., DNAM-1, a novel adhesion molecule involved in the cytolytic function of T lymphocytes, Immunity 4, 573-581 (1996)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		567888|DNAM-1	2.2E-112	[Homo sapiens][Adhesin/agglutinin][Plasma membrane] Adhesion glycoprotein, member of the immunoglobulin superfamily, involved in cell adhesion during cytotoxic T lymphocyte-mediated cytotoxicity and in the beta2 integrin LFA-1 (ITGAL) signaling pathway
				Shibuya, K. et al., Physical and functional association of LFA-1 with DNAM-1 adhesion molecule., Immunity 11, 615-23. (1999).
15	7521901CD1	g13021810	1.9E-63	[Homo sapiens] NK cell receptor
				Boles, K. S. et al., Molecular cloning of CS1, a novel human natural killer cell receptor belonging to the CD2 subset of the immunoglobulin superfamily, Immunogenetics 52, 302-307 (2001)
		789879|CRACC	1.3E-64	[Homo sapiens] Protein with low similarity to cluster of differentiation antigen 84 (human CD84), which binds the X-linked lymphoproliferative disease gene product SH2D1A (human SAP) and functions in the immune response, cell adhesion, and possibly leukocyte activation
16	7522003CD1	g13183883	1.6E-116	[Homo sapiens] PD-1-ligand 2 protein
				Latchman, Y. et al., PD-L2 is a second ligand for PD-1 and inhibits T cell activation, Nat. Immunol. 2, 261-268 (2001)
		692034|PDL2	1.2E-117	[Homo sapiens] Programmed death ligand 2, a member of B7 family, ligand for programmed death-1 (PDCD1) and may inhibit TCR- and CD28-mediated proliferation and cytokine production through cell cycle arrest
				Latchman, Y. et al., PD-L2 is a second ligand for PD-I and inhibits T cell activation., Nat Immunol 2, 261-8. (2001).
		618720|Btdc	3.7E-78	[Mus musculus] Programmed death ligand 2, a member of B7 family, ligand for programmed death-1 that inhibits T cell activation, involved in cell-mediated immune response and cell cycle arrest
				Tseng, S. Y. et al., B7-DC, a new dendritic cell molecule with potent costimulatory properties for T cells., J Exp Med 193, 839-46. (2001).
17	7522014CD1	g545773	3.8E-24	[Homo sapiens] surface antigen CD70; ligand for CD27
				Bowman, M. R. et al., The cloning of CD70 and its identification as the ligand for CD27, J. Immunol. 152, 1756-1761 (1994)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		338614|TNFSF7	2.8E-25	[Homo sapiens][Ligand][Plasma membrane] Tumor necrosis factor (ligand) superfamily member 7, transmembrane protein that is a ligand for the receptor CD27 (TNFRSF7), involved in B cell differentiation and antiapoptosis processes, induces activated T-cells to proliferate and become cytolytic
				Erlichman, B. et al., CD27 signals through PKC in human B cell lymphomas., Cytokine 11, 476-84. (1999).
18	7522038CD1	g2352941	2.3E-42	[Homo sapiens] leukocyte-associated Ig-like receptor-1
				Meyaard, L. et al., LAIR-1, a novel inhibitory receptor expressed on human mononuclear leukocytes, Immunity 7, 283-290 (1997)
		623778|LAIR1	1.7E-43	[Homo sapiens][Receptor (signaling)][Plasma membrane] Leukocyte-associated Ig-like receptor 1, inhibitory receptor that regulates B, T and NK cell activation, may be regulated by SHP-1 (PTPN6), sequesters NF-kB in the cytoplasm and induces programmed cell death; potential target for anti-leukemic therapy
				Meyaard, L. et al., Leukocyte-associated Ig-like receptor-1 functions as an inhibitory receptor on cytotoxic T cells., J Immunol 162, 5800-4 (1999).
19	7523429CD1	g292390	6.7E-11	[Homo sapiens] platelet factor 4
				Green, C. J. et al., Identification and characterization of PF4varl, a human gene variant of platelet factor 4, Mol. Cell. Biol. 9, 1445-1451 (1989)
		344702|PF4V1	4.9E-12	[Homo sapiens][Ligand][Extracellular (excluding cell wall)] Platelet factor 4 variant 1, a putative heparin-binding chemokine with a hydrophobic leader sequence and disulfide-bonded core, member of the platelet factor 4 family of heparin-binding proteins that are secreted during inflammatory responses
				Modi, W. S. et al., Localization of the human CXC chemokine subfamily on the long arm of chromosome 4 using radiation hybrids., Genomics 47, 136-9 (1998).
20	7523941CD1	g1864005	1.1E-65	[Homo sapiens] transmembrane protein
				Yoshiyama, K. et al., CD156 (human ADAM8): expression, primary amino acid sequence, and gene location., Genomics 41, 56-62 (1997).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		339052| ADAM8	8.0E-67	[Homo sapiens][Hydrolase;Protease (other than proteasomal)][Unspecified membrane;Plasma membrane] Disintegrin and metalloproteinase domain 8, a member of the ADAM family of disintegrin domain-containing zinc metalloproteases
				Yoshiyama, K. et al., CD156 (human ADAM8): expression, primary amino acid sequence, and gene location., Genomics 41, 56-62 (1997).
		583551| Adam8	1.4E-43	[Mus musculus][Hydrolase;Protease (other than proteasomal)][Plasma membrane] Disintegrin and metalloproteinase domain 8, a member of the ADAM family of disintegrin domain-containing zinc metalloproteases that plays a role in osteoclast formation and cell adhesion during neurodegeneration
				Amour, A. et al., The enzymatic activity of ADAM8 and ADAM9 is not regulated by TIMPs., FEBS Lett 524, 154-8. (2002).
21	7524607CD1	g16755850	1.2E-54	[Homo sapiens] parotid secretory protein
		761647|C20orf70	8.4E-56	[Homo sapiens] Protein with low similarity to parotid secretory protein (mouse *Psp*), *which* is a bacteria-binding protein that may function as an innate antimicrobial agent and has aberrant expression in nonobese diabetic mice
		751684|Psp	2.0E-15	[Rattus norvegicus][Extracellular (excluding cell wall)] Parotid secretory protein (submandibular gland protein A), a major secreted product of the submandibular gland acinar-cell progenitors and parotid glands, may play a role in bacterial binding, expression may be downregulated in diabetes
				Mirels, L. et al., Characterization of the rat salivary-gland B1-immunoreactive proteins., Biochem J 330, 437-44 (1998).
22	7524690CD1	g179726	1.2E-255	[Homo sapiens] C9 complement protein
				DiScipio, R. G. et al., Nucleotide sequence of cDNA and derived amino acid sequence of human complement component C9, Proc. Natl. Acad. Sci. U.S.A. 81, 7298-7302 (1984)
		334410|C9	1.5E-261	[Homo sapiens][Plasma membrane] Complement component 9, a subunit of the membrane attack complex, part of the classic and alternative complement pathways, forms a transmembrane channel, inhibits apoptosis; mutations lead to increased meningococcal meningitis infections

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Shiver, J. W. et al., The ninth component of human complement (C9). Functional activity of the b fragment., J Biol Chem 261, 9629-36 (1986).
		759154|C9	1.4E-181	[Rattus norvegicus][Small molecule-binding protein] Protein with high similarity to complement component 9 (murine Mm.29095), which is a subunit of the membrane attack complex, contains a low-density lipoprotein receptor class A domain, and a type 1 thrombospondin domain
				Lang, T. J. et al., Activation of the alternative complement pathway and production of factor H by skeletal myotubes, J Neuroimmunol 44, 185-92 (1993).
23	7524733CD1	g33781	9.3E-62	[Homo sapiens] interleukin-2
				Devos, R. et al., Molecular cloning of human interleukin 2 cDNA and its expression in E. coli, Nucleic Acids Res. 11, 4307-4323 (1983)
		618366|IL2	4.6E-60	[Homo sapiens][Ligand][Golgi;Cytoplasmic;Extracellular (excluding cell wall)] Interleukin-2 (T-cell growth factor), a T-cell-derived cytokine that promotes activation and proliferation of lymphocytes, involved in the immune response, implicated in Sjorgen's syndrome, autoimmune hemolytic anemia, and multiple sclerosis
				Lowenthal, J. W. et al., Similarities between interleukin-2 receptor number and affinity on activated B and T lymphocytes., Nature 315, 669-72 (1985).
		757398|Il2	8.0E-39	[Rattus norvegicus][Ligand] Interleukin-2, a T-cell-derived cytokine that promotes growth of B and T cells, likely involved in the immune response; human IL2 is implicated in Sjorgen's syndrome, autoimmune hemolytic anemia, and multiple sclerosis
				Lowenthal, J. W. et al., Similarities between interleukin-2 receptor number and affinity on activated B and T lymphocytes, Nature 315, 669-72 (1985).
24	7522128CD1	g30949	1.6E-82	[Homo sapiens] pre-pro polypeptide (AA -22 to 163)
				Gelin, C. et al., The E2 antigen, a 32 kd glycoprotein involved in T-cell adhesion processes, is the MIC2 gene product, EMBO J. 8, 3253-3259 (1989)
		336412|MIC2	1.1E-83	[Homo sapiens][Adhesin/agglutinin][Cytoplasmic;Plasma membrane] T-cell surface glycoprotein E2 (CD99), a cell surface glycoprotein involved in T cell adhesion and apoptosis, activates MAP kinases, may be involved in Golgi-membrane transport, downregulated in EBV-associated Hodgkins disease

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Hahn, M. J. et al., Differential activation of MAP kinase family members triggered by CD99 engagement., FEBS Lett 470, 350-4. (2000).
		788635| Mic2l1	5.4E-13	[Rattus norvegicus] MIC2 like 1 (rhombex-40), a putative transmembrane protein, highly expressed in ventral medullary surface neurons compared with cerebral cortex neurons
				Shimokawa, N. et al., Molecular cloning of Rhombex-40 a transmembrane protein from the ventral medullary surface of the rat brain by differential display, Life Sci 66, 2183-91.
25	7522158CD1	g12859845	1.7E-120	[Mus musculus] NOVEL PROTEIN SIMILAR TO X. LAEVIS CORTICAL THYMOCYTE MARKER CTX
		569052| CTXL	3.3E-23	[Homo sapiens][Unspecified membrane;Plasma membrane] Cortical thymocyte receptor-like, a member of the immunoglobulin superfamily with one variable and one constant C2-type domain; Xenopus homolog CTX is a cortical thymocyte marker
				Chretien, I. et al., CTX, a Xenopus thymocyte receptor, defines a molecular family conserved throughout vertebrates., Eur J Immunol 28, 4094-104 (1998).
		614107|Ctxl	1.2E-20	[Mus musculus] Cortical thymocyte receptor-like, a member of the immunoglobulin superfamily with one variable and one constant C2-type domain; Xenopus homolog CTX is a cortical thymocyte marker
26	7524191CD1	g15528835	1.8E-99	[Homo sapiens] Fc receptor-like protein 3
				Davis, R. S. et al., Identification of a family of Fc receptor homologs with preferential B cell expression, Proc. Natl. Acad. Sci. U.S.A. 98, 9772-9777 (2001)
		752423 |FCRH3	1.3E-100	[Homo sapiens] Fc receptor-like protein 3, an immunoglobulin superfamily member with putative regulatory roles in B cell development
				Davis, R. S. et al., Identification of a family of Fc receptor homologs with preferential B cell expression, Proc Natl Acad Sci U S A 98, 9772-7 (2001).
		743068| IRTA2	1.6E-28	[Homo sapiens] Immunoglobulin superfamily receptor translocation associated 2, a putative immunoreceptor with possible roles in B cell development and lymphomagenesis; deregulation of the gene by translocation occurs in some B cell malignancies
				Hatzivassiliou, G. et al., IRTA1 and IRTA2, novel immunoglobulin superfamily receptors expressed in B cells and involved in chromosome 1q21 abnormalities in B cell malignancy, Immunity 14, 277-89 (2001).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
27	7525225CD1	g180044	1.7E-170	[Homo sapiens] thymocyte antigen
				Martin, L. H. et al., Structure and expression of the human thymocyte antigens CD1a, CD1b, and CD1c, Proc. Natl. Acad. Sci. U.S.A. 84, 9189-9193 (1987)
		356777| CD1A	7.6E-170	[Homo sapiens][Small molecule-binding protein][Plasma membrane] CD1A antigen, a member of the CD1 family of surface molecules, interacts with beta 2-microglobulin (B2M) to present lipid and glycolipid antigens to T lymphocytes, may contribute to inflammatory process associated with atherosclerotic plaques
				Sugita, M. et al., CD1c molecules broadly survey the endocytic system, Proc Natl Acad Sci U S A 97, 8445-50 (2000).
		334514|CD1B	1.1E-94	[Homo sapiens][Small molecule-binding protein][Endosome/Endosomal vesicles;Cytoplasmic;Plasma membrane] CD1B antigen b polypeptide, binds and presents lipid and glycolipid antigens to T cells, expressed as a beta 2-microglobulin (B2M)-associated heterodimer; may play a role in the development of multiple sclerosis and other autoimmune deseases
				Martin, L. H. et al., Structure and expression of the human thymocyte antigens CD1a, CD1b, and CD1c., Proc Natl Acad Sci U S A 84, 9189-93 (1987).
28	7513053CD1	g595475	5.9E-180	[Homo sapiens] hFcRn
				Story, C. M. et al., A major histocompatibility complex class I-like Fc receptor cloned from human placenta: possible role in transfer of immunoglobulin G from mother to fetus, J. Exp. Med. 180, 2377-2381 (1994)
		340538| FCGRT	4.6E-181	[Homo sapiens][Receptor (protein translocation)] Fc receptor IgG alpha chain transporter, IgG receptor with similarity to class I major histocompatibility complex antigens, binds maternal IgG and transports it across placenta and intestinal epithelium to provide humoral immunity to fetuses and neonates
				Antohe, F. et al., Expression of functionally active FcRn and the differentiated bidirectional transport of IgG in human placental endothelial cells, Hum Immunol 62, 93-105. (2001).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		584773|Fcgrt	4.5E-116	[Mus musculus][Receptor (protein translocation)][Plasma membrane] Fc receptor IgG alpha chain transporter, a receptor for IgG with similarity to class I major histocompatibility complex antigens, binds maternal IgG and transports it across the intestinal epithelium to provide humoral immunity to neonates
				Borvak, J. et al., Functional expression of the MHC class I-related receptor, FcRn, in endothelial cells of mice, Int Immunol 10, 1289-98 (1998).
29	7513086CD1	g4539529	2.5E-85	[Homo sapiens] CD1d antigen
				Calabi, F. et al., Two classes of CD1 genes, Eur. J. Immunol. 19, 285-292 (1989)
		334518| CD1D	1.9E-86	[Homo sapiens][Ligand][Plasma membrane] CD1D antigen, major histocompatibility complex class I-like molecule, presents glycolipid antigens and activates natural killer T cells, involved in cellular defense and antitumor responses
				Kim, H. S. et al., Biochemical characterization of CD1d expression in the absence of beta2-microglobulin, J Biol Chem 274, 9289-95 (1999).
		581003| Cd1d1	2.9E-53	[Mus musculus][Endosome/Endosomal vesicles;Cytoplasmic;Plasma membrane] Cd1d1 antigen, major histocompatibility complex class I-like molecule, presents glycolipid antigens and activates natural killer T cells, involved in cellular defense, antitumor, and autoimmune responses
				Joyce, S., Woods, A. S., Yewdell, J. W., Bennink, J. R., De Silva, A. D., Boesteanu, A., Balk, S. P., Cotter, R. J., and Brutkiewicz, R. R. Natural ligand of mouse CD1d1: cellular glycosylphosphatidylinositol, Science 279, 1541-4. (1998).
30	7513557CD1	g219539	5.7E-59	[Homo sapiens] CGM7
				Kuroki, M. et al., Molecular cloning of nonspecific cross-reacting antigens in human granulocytes, J. Biol. Chem. 266, 11810-11817 (1991)
		334640|CEACAM4	4.5E-60	[Homo sapiens][Plasma membrane;Unspecified membrane] Member of the CEA subgroup of the immunoglobulin superfamily
				Kuroki, M. et al., Molecular cloning of nonspecific cross-reacting antigens in human granulocytes, J Biol Chem 266, 11810-7 (1991).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		710029| Ceacam1	5.9E-12	[Mus musculus][Adhesin/agglutinin][Plasma membrane] CEA-related cell adhesion molecule 1, a cell adhesion and signal transduction molecule, serves as a hepatitis virus receptor, negatively regulates tumor cell growth, may play roles in epithelial cell renewal and differentiation and in the immune response
				Turbide, C. et al., A mouse carcinoembryonic antigen gene family member is a calcium-dependent cell adhesion molecule, J Biol Chem 266, 309-15 (1991).
31	7513718CD1	g35797	1.5E-131	[Homo sapiens] pentaxin
				Breviario, F. et al., Interleukin-1-inducible genes in endothelial cells. Cloning of a new gene related to C-reactive protein and serum amyloid P component, J. Biol. Chem. 267, 22190-22197 (1992)
		337436|PTX3	1.2E-132	[Homo sapiens] Pentaxin-related gene, a member of the pentaxin family of acute-phase proteins, may play roles in inflammation and the bacterial defense response, may limit autoimmune reactions during inflammation, an early indicator of acute myocardial infarction
				Bottazzi, B. et al., Multimer formation and ligand recognition by the long pentraxin PTX3. Similarities and differences with the short pentraxins C-reactive protein and serum amyloid P component, J Biol Chem 272, 32817-23 (1997).
		582677|Ptx3	2.2E-114	[Mus musculus] Pentaxin-related gene, a member of the pentaxin family of acute-phase proteins, plays a role in inflammation and may play a role in the bacterial defense response; expression of human PTX3 is an early indicator of acute myocardial infarction
				Polentarutti, N. et al., Inducible expression of the long pentraxin PTX3 in the central nervous system, J Neuroimmunol 106, 87-94 (2000).
32	7514003CD1	g13274520	5.7E-47	[Homo sapiens] complement-c1q tumor necrosis factor-related protein
		703719|CTRP3	4.5E-48	[Homo sapiens] Collagenous repeat-containing sequence of 26-kDa, protein with very strong similarity to murine Cors, which is a collagenous repeat-containing protein with probable roles in skeletal development
		704197|Cors	6.9E-43	[Mus musculus] Collagenous repeat-containing sequence of 26-kDa, a secreted protein that may play a role in skeletal development

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Maeda, T. et al., Molecular cloning and characterization of a novel gene, CORS26, encoding a putative secretory protein and its possible involvement in skeletal development, J Biol Chem 276, 3628-34 (2001).

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7522043CD1	256	signal_cleavage: M1-A22	SPSCAN
			Signal Peptide: M1-A21, M1-G24, M1-P25, M1-A22	HMMER
			ANTIGEN PRECURSOR SIGNAL IMMUNOGLOBULIN FOLD GLYCOPROTEIN T-CELL SURFACE CD2 TRANSMEMBRANE PD010953: G24-I209	BLAST_PRODOM
			Potential Phosphorylation Sites: S89 S116 S139 S183 S224 T163 T235 T244 Y103	MOTIFS
			Potential Glycosylation Sites: N98 N142 N148 N172 N176 N204	MOTIFS
2	7523539CD1	433	signal_cleavage: M1-A30	SPSCAN
			Signal Peptide: M5-A30, M1-A30, M5-A31	HMMER
			BPI/LBP/CETP N-terminal domain: R39-H210	HMMER_SMART
			BPI/LBP/CETP C-terminal domain: A225-A428	HMMER_SMART
			LBP / BPI / CETP family, N-terminal domain: V32-T193	HMMER_PFAM
			LBP / BPI / CETP family, C-terminal domain: E195-Y432	HMMER_PFAM
			Cytosolic domain: M1-W15 Transmembrane domain: A16-V38 Non-cytosolic domain: R39-K433	TMHMMER
			Lipid-binding serum glycoprotein IPB001124: P34-K61, G78-A124, L163-Y206, H227-P263	BLIMPS_BLOCKS
			LBP / BPI / CETP family signature: R14-S86	PROFILESCAN
			PROTEIN PRECURSOR SIGNAL GLYCOPROTEIN LIPID TRANSPORT ANTIBIOTIC TRANSMEMBRANE LIPOPOLYSACCHARIDE-BINDING LBP PD006440: L18-E140 S110-K433	BLAST_PRODOM
			LIPOPOLYSACCHARIDE-BINDING PROTEIN DM02253 |P17213|11-486: I74-K433 |P17213|11-486: N11-E140	BLAST_DOMO
			LIPOPOLYSACCHARIDE-BINDING PROTEIN DM02253 |P17453|7-481: P95-Y432 |P17453|7-481: A12-E140	BLAST_DOMO
			LIPOPOLYSACCHARIDE-BINDING PROTEIN DM02253|P18428|5-474: L18-L131 W130-Y432	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			LIPOPOLYSACCHARIDE-BINDING PROTEIN DM02253|P17454|6-479: S17-Q164 L131-K433	BLAST_DOMO
			Potential Phosphorylation Sites: S41 S71 S119 S160 S349 T174 T192 T256 T271	MOTIFS
			Potential Glycosylation Sites: N326	MOTIFS
			Lbp_Bpi_Cetp: P34-P66	MOTIFS
3	7523587CD1	142	signal_cleavage: M1-A17	SPSCAN
			Signal Peptide: M1-A17, M1-E19	HMMER
			Potential Phosphorylation Sites: S33 S43 T59 T75 T110 T123	MOTIFS
			Potential Glycosylation Sites: N80 N87	MOTIFS
4	7523622CD1	450	signal_cleavage: M1-A30	SPSCAN
			Signal Peptide: M5-A30, M1-A30, M5-A31	HMMER
			BPI/LBP/CETP N-terminal domain: R39-H264	HMMER_SMART
			BPI/LBP/CETP C-terminal domain: A279-A445	HMMER_SMART
			LBP / BPI / CETP family, N-terminal domain: V32-T247	HMMER_PFAM
			LBP / BPI / CETP family, C-terminal domain: E249-Y449	HMMER_PFAM
			Cytosolic domain: M1-W15 Transmembrane domain: V16-V38 Non-cytosolic domain: R39-K450	TMHMMER
			Lipid-binding serum glycoprotein IPB001124: P34-K61, G78-A124, L217-Y260, H281-P317	BLIMPS_BLOCKS
			LBP / BPI / CETP family signature: R14-S86	PROFILESCAN
			PROTEIN PRECURSOR SIGNAL GLYCOPROTEIN LIPID TRANSPORT ANTIBIOTIC TRANSMEMBRANE LIPOPOLYSACCHARIDE-BINDING LBP PD006440: V16-M391 V392-K450	BLAST_PRODOM
			LIPOPOLYSACCHARIDE-BINDING PROTEIN DM02253|P17213|11-486: N11-M391 V392-K450	BLAST_DOMO
			LIPOPOLYSACCHARIDE-BINDING PROTEIN DM02253|P17453|7-481: R14-M391 V392-Y449	BLAST_DOMO
			LIPOPOLYSACCHARIDE-BINDING PROTEIN DM02253|P18428|5-474: L18-M391 E393-Y449	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			LIPOPOLYSACCHARIDE-BINDING PROTEIN DM02253|P17454|6-479: V16-M391 V392-K450	BLAST_DOMO
			Potential Phosphorylation Sites: S41 S71 S119 S156 T12 T228 T246 T310 T325	MOTIFS
			Potential Glycosylation Sites: N380	MOTIFS
			Lbp_Bpi_Cetp: P34-P66	MOTIFS
5	7523711CD1	84	signal_cleavage: M1-A17	SPSCAN
			Signal Peptide: M1-A17, M1-E19	HMMER
			Potential Phosphorylation Sites: S33 S43 T59	MOTIFS
6	7523729CD1	82	signal_cleavage: M1-P32	SPSCAN
			Signal Peptide: M1-P32	HMMER
			CLASS I CYTOKINE RECEPTOR PD043640: T30-R77	BLAST_PRODOM
			CLASS I CYTOKINE RECEPTOR PD145804: M1-R29	BLAST_PRODOM
			Potential Phosphorylation Sites: S70	MOTIFS
			Potential Glycosylation Sites: N51	MOTIFS
			Receptor_Cytokines_1: C41-W54	MOTIFS
7	7523763CD1	120	signal_cleavage: M1-S27	SPSCAN
			Signal Peptide: P9-S27, Q7-S27, Q7-P29, M1-G26, M1-S27, G5-S27	HMMER
			PROPERDIN PRECURSOR SIGNAL COMPLEMENT ALTERNATE PATHWAY GLYCOPROTEIN REPEAT DISEASE MUTATION PD012025: A24-P77	BLAST_PRODOM
			Potential Phosphorylation Sites: S40 S52 T25 T34	MOTIFS
8	7523006CD1	411	signal_cleavage: M1-S27	SPSCAN
			Signal Peptide: P9-S27	HMMER
			Signal Peptide: Q7-S27	HMMER
			Signal Peptide: Q7-P29	HMMER
			Signal Peptide: M1-G26	HMMER
			Signal Peptide: M1-S27	HMMER
			Signal Peptide: G5-S27	HMMER
			Thrombospondin type 1 repeats: W139-P191, W196-P255, W80-P134	HMMER_SMART
			Thrombospondin type 1 domain: S140-C190, S81-C133, A197-C254	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			PRECURSOR GLYCOPROTEIN SIGNAL REPEAT PD01719: W139-P166, R247-C254	BLIMPS_PRODOM
			PROPERDIN PRECURSOR SIGNAL COMPLEMENT ALTERNATE PATHWAY GLYCOPROTEIN REPEAT DISEASE MUTATION PD012461: P358-E410	BLAST_PRODOM
			PROPERDIN PRECURSOR SIGNAL COMPLEMENT ALTERNATE PATHWAY GLYCOPROTEIN REPEAT DISEASE MUTATION PD012025: A24-W80	BLAST_PRODOM
			PRECURSOR SIGNAL PROTEIN PROPERDIN SEMAPHORIN F G COMPLEMENT ALTERNATE PATHWAY PD017722: P252-P357, D129-C209, C72-C170	BLAST_PRODOM
			PROTEIN SIGNAL PRECURSOR REPEAT GLYCOPROTEIN THROMBOSPONDIN CIRCUMSPOROZOITE CELL ADHESION MALARIA PD000485: W139-C190	BLAST_PRODOM
			THROMBOSPONDIN TYPE 1 REPEAT DM00275|P27918|178-239: A178-G240	BLAST_DOMO
			THROMBOSPONDIN TYPE 1 REPEAT DM00275|P27918|119-176: T119-Q177	BLAST_DOMO
			THROMBOSPONDIN TYPE 1 REPEAT DM00275|P27918|365-419: Q307-M362	BLAST_DOMO
			THROMBOSPONDIN TYPE 1 REPEAT DM00275|P27918|300-363: C248-Q306	BLAST_DOMO
			Potential Phosphorylation Sites: S40 S52 S77 S112 S221 S225 S230 S262 S278 S323 S361 T25 T34 T92 T158 T249 T291	MOTIFS
			Potential Glycosylation Sites: N370	MOTIFS
			Receptor_Cytokines_2: G322-S328	MOTIFS
9	7523261CD1	208	signal_cleavage: M1-C28	SPSCAN
			Signal Peptide: M1-G24	HMMER
			Signal Peptide: M1-G24	HMMER
			Signal Peptide: M1-C28	HMMER
			Signal Peptide: M1-S30	HMMER
			C2 type Ig domains from SCOP: H21-P180	HMMER_INCY
			Cytosolic domain: S30-R173 Transmembrane domains: A10-F29, Y174-L196 Non-cytosolic domains: M1-R9, K197-N208	TMHMMER
			PRECURSOR SIGNAL CELL ADHESION LFA3 LFA3DELTA LYMPHOCYTE FUNCTION-ASSOCIATED ANTIGEN CD58 PD151527: M1-E122	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			LYMPHOCYTE FUNCTIONASSOCIATED ANTIGEN 3 PRECURSOR CD58 SURFACE GLYCOPROTEIN LFA3 IMMUNOGLOBULIN FOLD TRANSMEMBRANE SIGNAL GPI-ANCHOR ALTERNATIVE SPLICING PD044391: P125-C166	BLAST_PRODOM
			Potential Phosphorylation Sites: S76 S97 S98 S150 S170 T96 T113	MOTIFS
			Potential Glycosylation Sites: N40 N94 N109 N135	MOTIFS
10	7523277CD1	54	signal_cleavage: M1-D29	SPSCAN
			Cytosolic domain: E25-L54 Transmembrane domain: F5-V24 Non-cytosolic domain: M1-T4	TMHMMER
			Rhesus blood group protein signature PR00342: T4-L22	BLIMPS_PRINTS
			GLYCOPROTEIN RH50 ERYTHROCYTE MEMBRANE PROTEIN MUTANT RHESUS BLOOD GROUP-ASSOCIATED PLASMA PD150281: M1-E49	BLAST_PRODOM
			BLOOD; RH; GROUP; POLYPEPTIDE; DM07103|Q02094|1-408: M1-P52	BLAST_DOMO
			Potential Phosphorylation Sites: T31	MOTIFS
			Potential Glycosylation Sites: N37	MOTIFS
11	7523279CD1	174	signal_cleavage: M1-A37	SPSCAN
			Signal Peptide: T17-A37	HMMER
			Signal Peptide: M1-A37	HMMER
			Signal Peptide: L15-A37	HMMER
			Immunoglobulin: P43-I152	HMMER_SMART
			Immunoglobulin V-Type: I53-V131	HMMER_SMART
			Ig superfamily from SCOP: H39-K157	HMMER_INCY
			Cytosolic domain: M1-T23 Transmembrane domain: L24-V46 Non-cytosolic domain: L47-M174	TMHMMER
			GLYCOPROTEIN PRECURSOR IMMUNOGLOBULIN FOL. PD02886: C35-C85, C103-I152	BLIMPS_PRODOM
			PRECURSOR SIGNAL GLYCOPROTEIN IMMUNOGLOBULIN FOLD T-CELL TRANSMEMBRANE T-CELL-SPECIFIC SURFACE CD28 PD005676: C58-Y139	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			PRECURSOR SIGNAL CYTOTOXIC T-LYMPHOCYTE PROTEIN CTLA4 IMMUNOGLOBULIN FOLD T-CELL TRANSMEMBRANE PD012955: M1-F56	BLAST_PRODOM
			T-CELL SURFACE GLYCOPROTEIN CD28 DM03346|P16410|1-222: M1-P158	BLAST_DOMO
			T-CELL SURFACE GLYCOPROTEIN CD28 DM03346|P31043|1-220: V40-V151	BLAST_DOMO
			Potential Phosphorylation Sites: S49 S159 T96 Y127	MOTIFS
			Potential Glycosylation Sites: N113 N145	MOTIFS
12	7523296CD1	254	signal_cleavage: M1-W45	SPSCAN
			Signal Peptide: H26-A46	HMMER
			C-type lectin (CTL) or carbohydrate-recognition domain (CRD): C126-E249	HMMER_SMART
			Lectin C-type domain: P143-K250	HMMER_PFAM
			Cytosolic domain: M1-P31 Transmembrane domain: L32-S54 Non-cytosolic domain: K55-C254	TMHMMER
			C-type lectin domain signature and profile: S205-R251	PROFILESCAN
			Type II antifreeze protein signature PR00356: P125-C137, C137-C154, A155-F172, F208-A219, W234-I247	BLIMPS_PRINTS
			C-TYPE LECTIN DM00035|P20693|179-304: Q119-C248	BLAST_DOMO
			C-TYPE LECTIN DM00035|P06734|156-281: C123-C248	BLAST_DOMO
			C-TYPE LECTIN DM00035|A46274|248-377: C123-E249	BLAST_DOMO
			C-TYPE LECTIN DM00035|P02707|74-202: P125-K250	BLAST_DOMO
			Potential Phosphorylation Sites: S12 S28 S241 T3 T65 T66	MOTIFS
			Potential Glycosylation Sites: N120	MOTIFS
			Leucine_Zipper: L63-L84	MOTIFS
			C_Type_Lectin_1: C225-C248	MOTIFS
13	7521779CD1	303	signal_cleavage: M1-S25	SPSCAN
			Signal Peptide: M1-G21	HMMER
			Signal Peptide: M1-S25	HMMER
			Signal Peptide: M1-S24	HMMER
			RECEPTOR CHAIN PRECURSOR TRANSMEMBRANE. PD02382: C155-N170	BLIMPS_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			RECEPTOR INTERLEUKIN13 ALPHA2 CHAIN PRECURSOR BINDING PROTEIN TRANSMEMBRANE GLYCOPROTEIN SIGNAL PD128978: M1-L109	BLAST_PRODOM
			RECEPTOR PRECURSOR SIGNAL TRANSMEMBRANE GLYCOPROTEIN CHAIN ALPHA INTERLEUKIN5 ANTIGEN ALTERNATIVE PD008667: C113-I235	BLAST_PRODOM
			GROWTH FACTOR AND CYTOKINES RECEPTORS FAMILY DM02361|Q01344|142-354: L153-I276	BLAST_DOMO
			Potential Phosphorylation Sites: S24 S26 S122 S156 S250 T80 T287	MOTIFS
			Potential Glycosylation Sites: N115 N215	MOTIFS
14	7521826CD1	224	Signal Peptide: M1-A20	HMMER
			Immunoglobulin: N25-A129	HMMER_SMART
			Immunoglobulin C-2 Type: E31-G117	HMMER_SMART
			Cytosolic domain: N164-V224 Transmembrane domain: V141-L163 Non-cytosolic domain: M1-F140	TMHMMER
			PLATELET T CELL ACTIVATION ANTIGEN DNAM1 PD152553: Q41-V224	BLAST_PRODOM
			Potential Phosphorylation Sites: S115 S201 S217 T76 T101 T127 T181 T205 Y66	MOTIFS
			Potential Glycosylation Sites: N35 N74 N86 N119 N199	MOTIFS
15	7521901CD1	165	signal_cleavage: M1-A22	SPSCAN
			Signal Peptide: M1-A21	HMMER
			Signal Peptide: M1-A22	HMMER
			Signal Peptide: M1-G24	HMMER
			Signal Peptide: M1-P25	HMMER
			Potential Phosphorylation Sites: S89 S116 S133 T144 T153 Y103	MOTIFS
			Potential Glycosylation Sites: N98	MOTIFS
16	7522003CD1	228	signal_cleavage: M1-T24	SPSCAN
			Signal Peptide: M1-A19	HMMER
			Immunoglobulin: P26-K120	HMMER_SMART
			Ig superfamily from SCOP: V23-S122, R124-E201	HMMER_INCY
			b7/CD80/CD86 multiple Ig domain p: V23-R222	HMMER_INCY

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Potential Phosphorylation Sites: S122 T74 T116 T212 T213 T219 T220 Y174	MOTIFS
			Potential Glycosylation Sites: N37 N64 N157 N163 N189 N211	MOTIFS
17	7522014CD1	98	signal_cleavage: M1-A44	SPSCAN
			Signal Peptide: Y15-A44	HMMER
			Cytosolic domain: I38-R98 Transmembrane domain: Y15-C37 Non-cytosolic domain: M1-P14	TMHMMER
			CD27 LIGAND CD27L CD70 ANTIGEN CYTOKINE TRANSMEMBRANE GLYCOPROTEIN SIGNAL ANCHOR PD169505: M1-G64	BLAST_PRODOM
			Potential Phosphorylation Sites: S9 S52 S74 T63 T96	MOTIFS
18	7522038CD1	122	signal_cleavage: M1-T21	SPSCAN
			Signal Peptide: M1-A16	HMMER
			Signal Peptide: M1-T18	HMMER
			Signal Peptide: M1-T21	HMMER
			Signal Peptide: M1-E23	HMMER
			Potential Phosphorylation Sites: S78 S80 T21 T57 T67 T97	MOTIFS
			Potential Glycosylation Sites: N69	MOTIFS
19	7523429CD1	60	signal_cleavage: M1-A30	SPSCAN
			Signal Peptide: M17-A32	HMMER
			Signal Peptide: M17-G34	HMMER
			Signal Peptide: M17-S36	HMMER
			Signal Peptide: M1-G34	HMMER
			Signal Peptide: R11-G34	HMMER
			Cytosolic domain: R33-L60 Transmembrane domain: Q15-A32 Non-cytosolic domain: M1-R14	TMHMMER
			PLATELET FACTOR 4 VARIANT PRECURSOR PF4VAR1 PROTEOGLYCAN HEPARIN-BINDING SIGNAL PD055150: M1-S36	BLAST_PRODOM
			Potential Phosphorylation Sites: S44 S50 T13	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
20	7523941CD1	139	Signal Peptide: M1-S20, M1-A18, M1-P22	HMMER
			signal_cleavage: M1-S20	SPSCAN
			Reprolysin family propeptide: H71-G139	HMMER_PFAM
			DOMAIN MDC TRANSMEMBRANE METALLOPROTEINASE EGF-LIKE DISINTEGRIN-LIKE PRECURSOR CYSTEINE-RICH PD000935: Y29-L127	BLAST_PRODOM
			ZINC; NEUTRAL; METALLOPEPTIDASE; HEMORRHAGIC; DM00533|Q05910|1-187:L4-R128 DM00533|P15167|1-188:M26-R128 DM00533|P34182|1-185:M26-R128 DM00533|JC4342|1-188:M26-R128	BLAST_DOMO
			Potential Phosphorylation Sites: T133, Y107	MOTIFS
			Potential Glycosylation Sites: N67, N91	MOTIFS
21	7524607CD1	213	Signal Peptide: M1-E19, M1-S20, M1-T15	HMMER
			signal_cleavage: M1-S18	SPSCAN
			PAROTID PRECURSOR SECRETORY SIGNAL GLAND SALIVARY EBNER VON SUBMANDIBULAR MINOR PD011295: E19-I188	BLAST_PRODOM
			PAROTID SECRETORY PROTEIN DM04779|P07743|12-234:V13-N202 DM04779|B42337|12-235:V13-N202 DM04779|A42337|12-206:G12-Q204	BLAST_DOMO
			Potential Phosphorylation Sites: S20, S148, T50, T95, T167	MOTIFS
22	7524690CD1	474	Signal Peptide: M1-A21	HMMER
			signal_cleavage: M1-A21	SPSCAN
			MAC/Perforin domain: P138-L394	HMMER_PFAM
			Low-density lipoprotein receptor domain class A: D99-S136	HMMER_PFAM
			Thrombospondin type 1 domain: S46-C94	HMMER_PFAM
			Low-density lipoprotein receptor domain class A: D100-S136	HMMER_SMART
			membrane-attack complex / perforin: M208-N418	HMMER_SMART
			Thrombospondin type 1 repeats: M45-E95	HMMER_SMART

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Membrane attack complex components/perforin/complement C9 IPB001862: P56-N74, F234-V287, R373-V395, N408-E455	BLIMPS_BLOCKS
			Low density lipoprotein (LDL)-receptor class A (LDLRA) domain IPB002172: C112-E131	BLIMPS_BLOCKS
			Membrane attack complex components / perforin signature: L230-S281	PROFILESCAN
			Low density lipoprotein (LDL) receptor signature PR00261: G110-E131	BLIMPS_PRINTS
			Complement C9 signature PR00764: F62-C78, F105-C125, K239-Y261, N372-I391, A402-S421, C425-C445	BLIMPS_PRINTS
			COMPLEMENT PATHWAY GLYCOPROTEIN COMPONENT DOMAIN PRECURSOR EGF-LIKE SIGNAL CYTOLYSIS TRANSMEMBRANE PD004411:S136-K207 E205-N472	BLAST_PRODOM
			COMPLEMENT COMPONENT PRECURSOR GLYCOPROTEIN SIGNAL PATHWAY ALTERNATE PLASMA MEMBRANE ATTACK PD149976: S35-E98	BLAST_PRODOM
			MEMBRANE ATTACK COMPLEX COMPONENTS / PERFORIN DM01968|P02748|30-110:E30-R111	BLAST_DOMO
			MEMBRANE ATTACK COMPLEX COMPONENTS / PERFORIN DM01986|P02748|149-499:E205-D415 DM01986|P02748|149-499:E149-E249 DM01986|P48747|177-509:F176-H217 E205-D415 DM01986|P48770|176-499:E205-D415	BLAST_DOMO
			Potential Phosphorylation Sites: S39, S52, S129, S281, S376, S385, S421, T25, T79, T109, T170, T233, T243, T348, T363	MOTIFS
			Potential Glycosylation Sites: N330	MOTIFS
			LDL-receptor class A (LDLRA) domain signature: C112-C134	MOTIFS
			Membrane attack complex components / perforin signature: Y250-Y261	MOTIFS
23	7524733CD1	133	Signal Peptide: M1-S20, M4-A21, M1-S25, M4-S25, M4-S20, M1-A15	HMMER
			signal_cleavage: M1-S20	SPSCAN
			Interleukin 2: L7-S130	HMMER_PFAM
			Interleukin-2 family: M1-T133	HMMER_SMART
			Interleukin-2 IPB000779: M1-Q42, I86-T133	BLIMPS_BLOCKS
			Interleukin-2 signature: L38-H79	PROFILESCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Interleukin-2 signature PR00265: Y2-T30, A50-N77, N88-D109, T111-L132	BLIMPS_PRINTS
			T-CELL GROWTH FACTOR PRECURSOR IL-2 SIGNAL INTERLEUKIN-2 IMMUNE CYTOKINE TCGF PD003649:L7-N49 A50-S130	BLAST_PRODOM
			INTERLEUKIN-2 DM01614|P46649|1-153:M1-N49 A50-T133 DM01614|Q07885|1-153:M1-E52 A50-T133 DM01614|P26891|1-153:M1-E52 A50-T133 DM01614|Q08081|1-154:M1-N49 A50-T133	BLAST_DOMO
			Potential Phosphorylation Sites: S26, T27, T113	MOTIFS
			Potential Glycosylation Sites: N49	MOTIFS
			Interleukin-2 signature: T51-L63	MOTIFS
24	7522128CD1	218	Signal Peptide: M1-G25, M1-G24, M1-A20, M1-P22, M1-A31	HMMER
			signal_cleavage: M1-A20	SPSCAN
			Cytosolic domains: M1-A6, Q149-G218 Transmembrane domains: L7-F26, V126-Y148 Non-cytosolic domain: D27-G125	TMHMMER
			T-CELL E2 GLYCOPROTEIN CD99 ANTIGEN PRECURSOR MIC2 SURFACE SPLICING SIGNAL PD026577: A21-D66	BLAST_PRODOM
			Potential Phosphorylation Sites: S48, S89, S91, S100, S113, S195, S200, T41	MOTIFS
			Potential Glycosylation Sites: N177	MOTIFS
25	7522158CD1	346	Signal Peptide: M1-V21, M1-V24, M1-S20, W6-V21	HMMER
			signal_cleavage: M1-V21	SPSCAN
			Immunoglobulin: D28-S188	HMMER_SMART
			Immunoglobulin C-2 Type: E111-G177	HMMER_SMART
			Cytosolic domain: A216-A346 Transmembrane domain: G193-F215 Non-cytosolic domain: M1-V192	TMHMMER
			CELL SURFACE A33 ANTIGEN PRECURSOR IMMUNOGLOBULIN FOLD LIPOPROTEIN PALMITATE GLYCOPROTEIN PD155626: G84-E253	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			PROCYCLIC ACIDIC REPETITIVE PROTEIN DM07225|P08469|1-113:E276-P329 DM07225|P08469|1-113:T269-P325	BLAST_DOMO
			Potential Phosphorylation Sites: S179, S188, S265, S315, S337, T25, T161, T231, T256, T269, T273, T313	MOTIFS
			Potential Glycosylation Sites: N32, N38, N159, N178	MOTIFS
26	7524191CD1	221	Signal Peptide: M1-S17, M1-E15	HMMER
			signal_cleavage: M1-A52	SPSCAN
			Immunoglobulin:P29-E96, Q105-Q186	HMMER_SMART
			RECEPTOR FC IMMUNOGLOBULIN AFFIN PD01270: P21-H60, I72-H108, D114-N142	BLIMPS_PRODOM
			RECEPTOR IMMUNOGLOBULIN GAMMA PRECURSOR AFFINITY GLYCOPROTEIN TRANSMEMBRANE FC-GAMMA IGG-BINDING PD002534: L5-K138	BLAST_PRODOM
			MYELIN-ASSOCIATED GLYCOPROTEIN DM00682|P12319|8-196:L2-V185 DM00682|S42209|1-172:V19-V185 DM00682|S40204|28-216:M1-E187 DM00682|P27645|19-208:L3-Q186	BLAST_DOMO
			Potential Phosphorylation Sites: S88, T177	MOTIFS
27	7525225CD1	306	Signal Peptide: M1-T18, M1-G16	HMMER
			signal_cleavage: M1-G16	SPSCAN
			Immunoglobulin C-Type: H218-Q288	HMMER_SMART
			Ig superfamily from SCOP: P204-G287	HMMER_INCY
			GLYCOPROTEIN ANTIGEN T-CELL PRECURSOR SURFACE SIGNAL TRANSMEMBRANE IMMUNOGLOBULIN DOMAIN FAMILY PD004615: P13-Q199	BLAST_PRODOM
			IMMUNOGLOBULIN DM00001|P06126|205-288:E205-D289	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			CLASS I HISTOCOMPATIBILITY ANTIGEN DM00083|P06126|2-195:L2-A196 DM00083|P29017|2-197:F3-K195 DM00083|S47246|2-196:L4-A196	BLAST_DOMO
			Potential Phosphorylation Sites: S39, S76, S123, S277, S283, T91, T247	MOTIFS
			Potential Glycosylation Sites: N37, N60, N74, N145	MOTIFS
28	7513053CD1	326	signal_cleavage: M1-G23	SPSCAN
			Signal Peptide: P8-G23	HMMER
			Signal Peptide: P6-G23	HMMER
			Signal Peptide: P4-G23	HMMER
			Signal Peptide: M1-G23	HMMER
			Signal Peptide: M1-A24	HMMER
			Signal Peptide: M1-L28	HMMER
			Signal Peptide: M1-S26	HMMER
			Immunoglobulin C-Type: F216-P285	HMMER_SMART
			Class I Histocompatibility antigen, domains: S26-L197	HMMER_PFAM
			Immunoglobulins and major histocompatibility complex IPB000495: V218-L240, Y273-L290	BLIMPS_BLOCKS
			MHC CLASS I ANTIGEN PRECURSOR SIGNAL CHAIN HISTOCOMPATIBILITY ALPHA GLYCOPROTEIN PD000050: S26-L197	BLAST_PRODOM
			CLASS I HISTOCOMPATIBILITY ANTIGEN DM00083|P13599|6-195: L11-R194	BLAST_DOMO
			CLASS I HISTOCOMPATIBILITY ANTIGEN DM00083|P15978|6-207: P6-P202	BLAST_DOMO
			CLASS I HISTOCOMPATIBILITY ANTIGEN DM00083|S33355|2-203: L15-P202	BLAST_DOMO
			IMMUNOGLOBULIN DM00001|P13599|204-286: P203-L286	BLAST_DOMO
			Immunoglobulins and major histocompatibility complex proteins signature: Y273-H279	MOTIFS
			Potential Phosphorylation Sites: S63 S161 S204 S267 S295 T130 T264 T307 Y273	MOTIFS
			Potential Glycosylation Sites: N125	MOTIFS
29	7513086CD1	158	signal_cleavage: M1-G17	SPSCAN
			Signal Peptide: M1-A15	HMMER
			Signal Peptide: M1-A19	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Signal Peptide: M1-E20	HMMER
			Signal Peptide: M1-Q23	HMMER
			Signal Peptide: M1-P22	HMMER
			PRECURSOR SIGNAL T CELL GLYCOPROTEIN SURFACE IMMUNOGLOBULIN FOLD ANTIGEN TRANSMEMBRANE MULTIGENE PD004615: G17-W158	BLAST_PRODOM
			CLASS I HISTOCOMPATIBILITY ANTIGEN DM00083|I49581|2-205: L4-W158	BLAST_DOMO
			CLASS I HISTOCOMPATIBILITY ANTIGEN DM00083|P11609|2-198: L4-W158	BLAST_DOMO
			CLASS I HISTOCOMPATIBILITY ANTIGEN DM00083|P15813|2-196: G2-W158	BLAST_DOMO
			CLASS I HISTOCOMPATIBILITY ANTIGEN DM00083|P29017|2-197: L4-T152	BLAST_DOMO
			Potential Phosphorylation Sites:S67 T42 T64 T75 T147	MOTIFS
			Potential Glycosylation Sites: N38 N60 N126	MOTIFS
30	7513557CD1	124	signal_cleavage: M1-A62	SPSCAN
			Cytosolic domain: S57-S124 Transmembrane domain: A34-L56 Non-cytosolic domain: M1-V33	TMHMMER
			ANTIGEN CARCINOEMBRYONIC CGM1 PRECURSOR CD66D IMMUNOGLOBULIN FOLD SIGNAL GLYCOPROTEIN TRANSMEMBRANE PD030936: R61-S124	BLAST_PRODOM
			CARCINOEMBRYONIC ANTIGEN PRECURSOR AMINO-TERMINAL DOMAIN DM02898|A40428|149-227: P29-S108	BLAST_DOMO
			CARCINOEMBRYONIC ANTIGEN PRECURSOR AMINO-TERMINAL DOMAIN DM02898|P40198|149-235: P29-L106	BLAST_DOMO
			Potential Phosphorylation Sites:S83 S94 T59	MOTIFS
31	7513718CD1	247	signal_cleavage: M1-A17	SPSCAN
			Signal Peptide: M1-A17	HMMER
			Signal Peptide: M1-N19	HMMER
			Pentaxin / C-reactive protein / pentaxin family: T41-S247	HMMER_SMART
			Pentaxin family: T47-S247	HMMER_PFAM
			Pentaxin family IPB001759: D43-I77, F91-Y105, H135-G153, I176-I214, G227-G241	BLIMPS_BLOCKS
			Pentaxin family signature: F113-M162	PROFILESCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Pentaxin signature PR00895: L70-V84, F91-Y105, H135-G153, M162-N181, C184-N203, N203-T217, H224-T235	BLIMPS_PRINTS
			PRECURSOR SIGNAL PENTAXIN PROTEIN GLYCOPROTEIN PLASMA C-REACTIVE CALCIUM ACUTE PHASE PD002153: L50-G242	BLAST_PRODOM
			PENTAXIN-RELATED PROTEIN PTX3 PRECURSOR TUMOR NECROSIS FACTOR INDUCIBLE TSG14 PENTAXIN PD037352: M1-D43	BLAST_PRODOM
			C-REACTIVE PROTEIN DM00835|P06205|8-241: E42-G242	BLAST_DOMO
			C-REACTIVE PROTEIN DM00835|P26022|139-380: C45-S247	BLAST_DOMO
			C-REACTIVE PROTEIN DM00835|P47970|187-426: I49-A243	BLAST_DOMO
			C-REACTIVE PROTEIN DM00835|P47971|194-431: N32-A243	BLAST_DOMO
			Pentaxin family signature: H135-S142	MOTIFS
			Potential Phosphorylation Sites:S55 S196 S210 T95 T158 Y105	MOTIFS
			Potential Glycosylation Sites: N86	MOTIFS
32	7514003CD1	106	signal_cleavage: M1-C22	SPSCAN
			Signal Peptide: M1-C22	HMMER
			Signal Peptide: M1-D24	HMMER
			Complement C1q protein IPB001073: G54-G78	BLIMPS_BLOCKS
			Potential Phosphorylation Sites:S48 T77	MOTIFS

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
33/7522043CB1/ 814	1-660, 1-814, 2-813, 209-389, 660-814
34/7523539CB1/ 1402	1-761, 6-1401, 18-37, 50-919, 72-895, 687-1402, 704-1402, 804-1402, 833-933
35/7523587CB1/ 1863	1-643, 1-866, 2-1862, 329-1244, 541-1252, 634-1340, 662-1551, 1024-1863, 1125-1863, 1129-1863
36/7523622CB1/ 1451	1-631, 1-738, 1-750, 2-1450, 17-36, 44-738, 46-750, 550-1288, 550-1451, 552-1404, 683-1451
37/7523711CB1/ 1685	1-637, 1-679, 2-1684, 626-1400, 627-1268, 848-1685, 946-1685, 950-1685
38/7523729CB1/ 1835	1-845, 2-1834, 178-1129, 277-1131, 631-1473, 637-1422, 914-1835, 1076-1834
39/ 7523763CB1/ 1495	1-763, 2-1494, 4-918, 189-904, 719-1364, 719-1495, 807-1495
40/ 7523006CB1/ 1313	1-809, 2-1312, 477-1313
41/ 7523261CB1/ 691	1-690, 1-691, 537-691
42/ 7523277CB1/ 954	1-848, 2-953, 137-954, 161-942, 161-954, 172-954
43/ 7523279CB1/ 591	1-591, 2-590
44/ 7523296CB1/ 766	1-579, 1-766, 2-765, 163-766, 168-766, 172-766, 226-766, 255-766, 265-766
45/ 7521779CB1/ 1091	1-1002, 393-1090, 611-1091
46/ 7521826CB1/ 703	1-703, 2-702

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
47/ 7521901CB1/ 541	1-541, 2-540, 209-541
48/ 7522003CB1/ 713	1-711, 1-713, 2-712
49/ 7522014CB1/ 648	1-648, 2-647, 5-648, 429-648
50/ 7522038CB1/ 378	1-377, 1-378, 271-377, 277-378
51/ 7523429CB1/ 676	1-477, 1-541, 1-551, 1-559, 1-563, 1-676, 9-665, 10-563
52/ 7523941CB1/ 2481	1-667, 2-612, 2-2480, 583-1469, 684-2480, 977-1855, 1091-1859, 1709-2481
53/ 7524607CB1/ 689	1-689, 6-688, 185-323, 321-475, 476-689
54/ 7524690CB1/ 1461	1-591, 1-869, 2-824, 671-1461
55/ 7524733CB1/ 457	1-457, 2-444
56/ 7522128CB1/ 1273	1-1273, 95-265, 95-271, 113-279, 113-996, 134-614, 381-1238
57/ 7522158CB1/ 1132	1-880, 10-1101, 256-1132
58/ 7524191CB1/ 2214	1-868, 2-1955, 673-1450, 673-1583, 788-1655, 809-1655, 885-1710, 920-1710, 1062-1171, 1062-1304, 1062-1315, 1341-1594, 1599-2214, 1812-2214
59/ 7525225CB1/ 965	1-828, 137-965, 168-964

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
60/ 7513053CB1/ 1829	1-641, 4-213, 4-1389, 7-125, 14-265, 18-118, 18-295, 20-351, 31-694, 58-530, 59-407, 60-614, 63-371, 63-400, 64-706, 80-626, 89-404, 91-374, 104-711, 140-385, 162-429, 166-417, 167-377, 167-409, 167-416, 167-420, 170-424, 170-431, 170-440, 170-441, 170-442, 170-449, 170-451, 170-458, 170-638, 170-669, 170-676, 170-690, 170-722, 170-747, 170-778, 170-976, 173-732, 175-966, 177-437, 178-416, 178-479, 179-444, 181-442, 181-775, 186-431, 187-492, 193-424, 193-451, 193-810, 194-453, 194-478, 195-485, 195-506, 196-459, 200-479, 217-754, 218-489, 218-500, 220-591, 221-590, 221-598, 225-447, 225-457, 228-489, 231-885, 237-856, 238-515, 240-508, 241-397, 241-515, 241-784, 241-808, 245-549, 250-505, 253-440, 254-428, 254-578, 259-417, 260-486, 260-560, 261-861, 280-599, 281-450, 281-543, 289-525, 289-556, 296-582, 297-565, 310-1028, 313-416, 316-485, 320-915, 321-697, 321-993, 330-595, 330-598, 332-607, 340-969, 343-969, 346-877, 348-1015, 352-645, 384-860, 390-648, 392-719, 392-1054, 395-657, 395-907, 396-645,
	396-798, 397-686, 401-624, 401-966, 405-688, 406-681, 408-689, 412-688, 424-1034, 426-509, 426-648, 426-670, 426-684, 427-692, 430-1054, 432-951, 432-977, 433-771, 435-712, 436-664, 438-610, 438-672, 438-690, 440-639, 442-693, 443-729, 447-669, 447-756, 447-827, 451-576, 452-695, 454-692, 454-719, 455-655, 460-714, 462-727, 462-936, 464-707, 466-746, 467-1015, 469-749, 471-697, 471-701, 472-931, 473-819, 474-744, 484-757, 490-703, 492-741, 502-714, 502-756, 503-791, 505-756, 505-761, 506-738, 508-970, 509-744, 518-767, 518-800, 520-966, 532-806, 538-1016, 539-763, 539-801, 550-777, 556-860, 557-796, 558-782, 558-800, 560-791, 560-820, 561-800, 561-807, 565-823, 569-741, 572-708, 573-827, 579-737, 593-846, 598-852, 600-772, 606-935, 608-774, 611-963, 612-840, 618-863, 620-842, 621-893, 622-895, 623-842, 626-837, 627-835, 630-825, 630-831, 632-973, 635-846, 641-916, 643-924, 644-916, 645-871, 645-943, 647-915, 656-943, 665-904, 665-907, 672-910, 675-930, 676-934, 676-941, 676-952, 678-953, 679-944, 680-916,
	681-948, 688-940, 690-946, 690-984, 693-931, 693-967, 702-971, 705-967, 711-958, 711-1011, 715-910, 719-966, 727-972, 728-966, 731-933, 732-986, 734-976, 735-960, 739-1025, 741-1009, 741-1012, 743-1038, 746-1007, 748-990, 751-964, 754-1038, 755-928, 757-1393, 765-1054, 766-1049, 772-1045, 777-1054, 782-1020, 784-967, 784-970, 784-983, 784-1006, 784-1019, 784-1292, 787-1027, 788-904, 789-1054, 798-1054, 799-1004, 800-1007, 802-1007, 808-976, 808-1021, 808-1054, 811-1054, 817-1010, 819-1034, 830-1054, 849-955, 896-1054, 1012-1325, 1054-1227, 1054-1247, 1054-1271, 1054-1281, 1054-1285, 1054-1286, 1054-1294, 1054-1303, 1054-1305, 1054-1313, 1054-1332, 1054-1368, 1054-1374, 1054-1376, 1054-1377, 1054-1380, 1054-1406, 1055-1375, 1056-1301, 1056-1359, 1057-1297, 1062-1333, 1065-1284, 1068-1314, 1069-1332, 1069-1377, 1075-1398, 1080-1285, 1083-1341, 1083-1345, 1083-1354, 1083-1377, 1084-1308, 1084-1346, 1084-1354, 1084-1360, 1085-1397, 1092-1378, 1105-1285, 1105-1350, 1108-1352, 1110-1325, 1110-1379, 1115-1390,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1119-1380, 1139-1377, 1149-1373, 1151-1394, 1157-1406, 1162-1345, 1162-1361, 1162-1392, 1164-1406, 1170-1456, 1172-1411, 1176-1407, 1189-1405, 1194-1396, 1201-1391, 1209-1316, 1215-1346, 1224-1374, 1225-1382, 1228-1386, 1235-1370, 1235-1381, 1235-1382, 1235-1383, 1235-1390, 1235-1391, 1235-1392, 1235-1396, 1235-1428, 1235-1442, 1235-1829, 1236-1375, 1236-1391, 1257-1354, 1257-1355, 1257-1370, 1257-1372, 1257-1373, 1257-1374, 1257-1375, 1257-1376, 1257-1377, 1257-1378, 1257-1381, 1257-1387, 1258-1373, 1258-1376, 1259-1374, 1260-1364, 1260-1375, 1261-1376, 1261-1377, 1261-1378, 1262-1373, 1262-1376, 1262-1377, 1263-1352, 1263-1366, 1263-1370, 1263-1373, 1263-1374, 1263-1375, 1263-1376, 1263-1377, 1274-1364, 1274-1377, 1288-1383, 1293-1472, 1295-1380, 1295-1432, 1298-1377
61/7513086CB1/ 1823	1-216, 3-197, 20-1795, 125-655, 126-394, 249-489, 249-741, 249-754, 249-775, 249-780, 249-812, 249-834, 249-851, 249-866, 249-905, 303-824, 383-555, 403-964, 458-788, 458-844, 497-643, 536-1083, 540-1240, 553-1028, 554-968, 556-900, 557-1093, 571-926, 589-1151, 591-1286, 593-923, 593-1198, 594-1140, 638-1106, 656-1281, 674-1200, 687-895, 698-1196, 709-1405, 718-1273, 733-1596, 762-1316, 762-1350, 765-1321, 778-1517, 779-1438, 789-1360, 803-1444, 840-1303, 876-1152, 905-1212, 908-1596, 946-1596, 949-1199, 949-1297, 950-1464, 963-1603, 974-1444, 987-1755, 1015-1273, 1126-1823, 1136-1414, 1215-1485, 1222-1747, 1247-1555, 1255-1752, 1259-1771, 1284-1823, 1285-1454, 1285-1488, 1285-1570, 1285-1696, 1295-1823, 1309-1731, 1309-1761, 1351-1562, 1359-1811, 1362-1823, 1368-1812, 1369-1823, 1370-1809, 1372-1640, 1375-1810, 1396-1675, 1407-1622, 1429-1813, 1431-1809, 1434-1823, 1435-1809, 1435-1823, 1440-1823, 1443-1811, 1451-1815, 1511-1763, 1523-1809, 1559-1809, 1568-1809, 1675-1823
62/7513557CB1/ 704	1-704, 4-704
63/7513718CB1/ 864	1-785, 48-864
64/7514003CB1/ 1566	1-710, 2-1566, 6-177

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
60	7513053CB1	BRSTTUT03
61	7513086CB1	LNODNOT11

Table 6

Library	Vector	Library Description
BRSTTUT03	PSPORT1	Library was constructed using RNA isolated from breast tumor tissue removed from a 58-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated multicentric invasive grade 4 lobular carcinoma. The mass was identified in the upper outer quadrant, and three separate nodules were found in the lower outer quadrant of the left breast. Patient history included skin cancer, rheumatic heart disease, osteoarthritis, and tuberculosis. Family history included cerebrovascular disease, coronary artery aneurysm, breast cancer, prostate cancer, atherosclerotic coronary artery disease, and type I diabetes.
LNODNOT11	pINCY	Library was constructed using RNA isolated from lymph node tissue removed from a 16-month-old Caucasian male who died from head trauma. Patient history included bronchitis.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value = 1.0E-8 or less; Full Length sequences: Probability value = 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value = 1.06E-6; Assembled ESTs: fasta Identity = 95% or greater and Match length = 200 bases or greater; fastx E value = 1.0E-8 or less; Full Length sequences: fastx score = 100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value = 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM, INCY, SMART or TIGRFAM hits: Probability value = 1.0E-3 or less; Signal peptide hits: Score = 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score ≥ GCG specified "HIGH" value for that particular Prosite motif. Generally, score = 1.4-2.1.

Table 7

Program	Description	Reference	Parameter Threshold
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score = 120 or greater; Match length = 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score = 3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
34	7523539	1489370T6	SNP00000213	243	1385	G	G	A	noncoding	n/a	n/a	n/a	n/a
34	7523539	1527490H1	SNP00014954	59	100	C	T	C	A16	n/a	n/a	n/a	n/a
34	7523539	1527490H1	SNP00066163	75	116	G	G	C	L21	n/a	n/a	n/a	n/a
34	7523539	1527490T6	SNP00000213	184	1386	A	G	A	noncoding	n/a	n/a	n/a	n/a
34	7523539	1619738T6	SNP00000213	229	1389	A	G	A	noncoding	n/a	n/a	n/a	n/a
35	7523587	3256488H1	SNP00071936	148	1033	C	C	T	noncoding	n/a	n/a	n/a	n/a
35	7523587	4516239F6	SNP00071936	270	1035	C	C	T	noncoding	n/a	n/a	n/a	n/a
35	7523587	4516239F6	SNP00154635	411	1176	T	C	T	noncoding	n/a	n/a	n/a	n/a
35	7523587	6454052F7	SNP00127785	173	177	G	G	A	P45	n/a	n/a	n/a	n/a
35	7523587	7345137H1	SNP00154635	563	1174	C	C	T	noncoding	n/a	n/a	n/a	n/a
35	7523587	7709976J1	SNP00071936	248	1034	C	C	T	noncoding	n/a	n/a	n/a	n/a
35	7523587	7709976J1	SNP00154635	106	1175	C	C	T	noncoding	n/a	n/a	n/a	n/a
36	7523622	1489370T6	SNP00000213	243	1434	G	G	A	noncoding	n/a	n/a	n/a	n/a
36	7523622	1527490H1	SNP00014954	59	98	C	T	C	A16	n/a	n/a	n/a	n/a
36	7523622	1527490H1	SNP00066163	75	114	G	G	C	L21	n/a	n/a	n/a	n/a
36	7523622	1527490T6	SNP00000213	184	1435	A	G	A	noncoding	n/a	n/a	n/a	n/a
36	7523622	1619738T6	SNP00000213	229	1438	A	G	A	noncoding	n/a	n/a	n/a	n/a
37	7523711	3256488H1	SNP00071936	148	854	C	C	T	noncoding	n/a	n/a	n/a	n/a
37	7523711	4516239F6	SNP00071936	270	856	C	C	T	noncoding	n/a	n/a	n/a	n/a
37	7523711	4516239F6	SNP00154635	411	997	T	C	T	noncoding	n/a	n/a	n/a	n/a
37	7523711	6454052F7	SNP00127785	173	178	G	G	A	P45	n/a	n/a	n/a	n/a
37	7523711	7345137H1	SNP00154635	563	995	C	C	T	noncoding	n/a	n/a	n/a	n/a
37	7523711	7709976J1	SNP00071936	248	855	C	C	T	noncoding	n/a	n/a	n/a	n/a
37	7523711	7709976J1	SNP00154635	106	996	C	C	T	noncoding	n/a	n/a	n/a	n/a
39	7523763	1562752T6	SNP00035104	80	1388	C	C	T	noncoding	n/a	n/a	n/a	n/a
39	7523763	1562752T6	SNP00071294	251	1217	C	C	A	noncoding	n/d	n/a	n/a	n/a
39	7523763	1880448T6	SNP00035103	242	1230	C	C	T	noncoding	n/a	n/a	n/a	n/a
39	7523763	1880448T6	SNP00071294	148	1325	C	C	A	noncoding	n/a	n/a	n/a	n/a
39	7523763	2464756H1	SNP00035104	49	1343	C	C	T	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
39	7523763	2757313H1	SNP00071294	30	1172	C	C	A	noncoding	n/d	n/a	n/a	n/a
39	7523763	2850837T6	SNP00035103	309	1083	C	C	T	noncoding	n/a	n/a	n/a	n/a
39	7523763	2850837T6	SNP00071294	216	1176	C	C	A	noncoding	n/d	n/a	n/a	n/a
39	7523763	4447844H1	SNP00035103	208	1079	C	C	T	noncoding	n/a	n/a	n/a	n/a
39	7523763	7733869J2	SNP00035103	444	1093	C	C	T	noncoding	n/a	n/a	n/a	n/a
39	7523763	7733869J2	SNP00035104	180	1357	C	C	T	noncoding	n/a	n/a	n/a	n/a
39	7523763	7733869J2	SNP00071294	351	1186	C	C	A	noncoding	n/d	n/a	n/a	n/a
40	7523006	1562752T6	SNP00035104	80	1206	C	C	T	H385	n/a	n/a	n/a	n/a
40	7523006	1562752T6	SNP00071294	251	1035	C	C	A	S328	n/d	n/a	n/a	n/a
40	7523006	1880448T6	SNP00035103	242	1048	C	C	T	R333	n/a	n/a	n/a	n/a
40	7523006	1880448T6	SNP00071294	148	1143	C	C	A	D364	n/a	n/a	n/a	n/a
40	7523006	2464756H1	SNP00035104	49	1161	C	C	T	N370	n/a	n/a	n/a	n/a
40	7523006	2757313H1	SNP00071294	30	990	C	C	A	Y313	n/d	n/a	n/a	n/a
40	7523006	2850837T6	SNP00035103	309	901	C	C	T	R284	n/a	n/a	n/a	n/a
40	7523006	2850837T6	SNP00071294	216	994	C	C	A	L315	n/d	n/a	n/a	n/a
40	7523006	4447844H1	SNP00035103	208	897	C	C	T	I282	n/a	n/a	n/a	n/a
40	7523006	7733869J2	SNP00035103	444	911	C	C	T	S287	n/a	n/a	n/a	n/a
40	7523006	7733869J2	SNP00035104	180	1175	C	C	T	A375	n/a	n/a	n/a	n/a
40	7523006	7733869J2	SNP00071294	351	1004	C	C	A	S318	n/d	n/a	n/a	n/a
41	7523261	6123457H1	SNP00066909	503	520	T	T	C	P162	n/d	n/d	n/d	n/d
48	7522003	3207811F6	SNP00145518	140	502	G	G	C	K163	n/a	n/a	n/a	n/a
48	7522003	3844031F8	SNP00145518	369	478	C	G	C	C155	n/a	n/a	n/a	n/a
49	7522014	1272144T6	SNP00018905	352	401	C	C	T	noncoding	n/a	n/a	n/a	n/a
49	7522014	1272144T6	SNP00058223	169	584	G	G	C	noncoding	0.99	n/a	n/a	n/a
49	7522014	2017463T6	SNP00058223	176	575	C	G	C	noncoding	0.99	n/a	n/a	n/a
49	7522014	219730H1	SNP00018905	9	391	C	C	T	noncoding	n/a	n/a	n/a	n/a
49	7522014	219730H1	SNP00058223	192	574	G	G	C	noncoding	0.99	n/a	n/a	n/a
50	7522038	1931275F6	SNP00020230	138	8	C	C	T	P3	n/a	n/a	n/a	n/a
52	7523941	1222454H1	SNP00041281	42	1461	G	G	A	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
52	7523941	1388967H1	SNP00010779	126	132	T	T	C	W35	n/a	n/a	n/a	n/a
52	7523941	2432550H1	SNP00047708	169	1073	G	A	G	noncoding	n/a	n/a	n/a	n/a
52	7523941	2432550H1	SNP00051866	64	968	T	T	C	noncoding	n/a	n/a	n/a	n/a
52	7523941	6528250H1	SNP00151570	199	1123	C	C	G	noncoding	n/a	n/a	n/a	n/a
53	7524607	1325602H1	SNP00105071	171	390	A	A	G	T124	n/d	n/d	n/d	n/d
53	7524607	1325602H1	SNP00105072	246	465	T	T	C	L149	n/a	n/a	n/a	n/a
53	7524607	1326305H1	SNP00137858	104	325	C	C	T	P102	n/a	n/a	n/a	n/a
53	7524607	1824780H1	SNP00105070	252	225	A	A	G	S69	n/d	n/d	n/d	n/d
53	7524607	1824780T6	SNP00105071	420	459	A	A	G	I147	n/a	n/a	n/a	n/a
53	7524607	1824780T6	SNP00111130	303	576	A	A	G	I186	n/d	n/d	n/d	n/d
53	7524607	1825090H1	SNP00111130	99	505	A	A	G	N162	n/d	n/d	n/d	n/d
53	7524607	1825123F6	SNP00105072	33	483	T	T	C	stop155	n/a	n/a	n/a	n/a
53	7524607	1825123F6	SNP00111130	73	523	A	A	G	Q168	n/d	n/d	n/d	n/d
53	7524607	1825123T6	SNP00105072	356	454	T	T	C	I145	n/a	n/a	n/a	n/a
53	7524607	1825123T6	SNP00111130	316	494	A	A	G	K158	n/d	n/d	n/d	n/d
54	7524690	085839H1	SNP00039727	112	110	G	G	C	S34	n/d	n/d	n/d	n/d
56	7522128	068440H1	SNP00003562	19	527	C	C	T	A121	n/a	n/a	n/a	n/a
56	7522128	1005401H1	SNP00141588	85	251	T	T	C	S29	n/a	n/a	n/a	n/a
56	7522128	1303089H1	SNP00096086	162	305	C	C	T	P47	n/a	n/a	n/a	n/a
56	7522128	1305754H1	SNP00138849	107	345	G	G	T	D61	n/a	n/a	n/a	n/a
56	7522128	1350913H1	SNP00023417	21	66	T	T	C	noncoding	n/a	n/a	n/a	n/a
56	7522128	1350913H1	SNP00144823	63	109	C	C	T	noncoding	n/a	n/a	n/a	n/a
56	7522128	1351738H1	SNP00096087	87	435	T	T	C	S91	n/a	n/a	n/a	n/a
56	7522128	1356937H1	SNP00023420	36	715	A	A	G	H184	0.88	0.95	0.91	0.91
56	7522128	1398562H1	SNP00023419	68	712	G	G	A	G183	n/a	n/a	n/a	n/a
56	7522128	1604845H1	SNP00136016	63	532	C	C	G	A123	n/a	n/a	n/a	n/a
56	7522128	1820467F6	SNP00096086	242	318	C	C	T	H52	n/a	n/a	n/a	n/a
56	7522128	1820467F6	SNP00138849	282	358	G	G	T	G65	n/a	n/a	n/a	n/a
56	7522128	1845464T6	SNP00018996	51	1167	T	T	C	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
56	7522128	1845464T6	SNP00092992	76	1142	C	C	T	noncoding	n/a	n/a	n/a	n/a
56	7522128	2061685R6	SNP00141588	102	255	T	T	C	S31	n/a	n/a	n/a	n/a
56	7522128	2113568H1	SNP00051500	9	75	A	A	G	noncoding	n/a	n/a	n/a	n/a
56	7522128	214323H1	SNP00018996	77	1165	T	T	C	noncoding	n/a	n/a	n/a	n/a
56	7522128	214323H1	SNP00092992	52	1140	C	C	T	noncoding	n/a	n/a	n/a	n/a
56	7522128	3369807H1	SNP00003561	173	251	T	T	C	S29	n/a	n/a	n/a	n/a
56	7522128	3908772H1	SNP00096087	44	437	T	T	C	S91	n/a	n/a	n/a	n/a
56	7522128	3909165H1	SNP00136016	141	534	C	C	G	P124	n/a	n/a	n/a	n/a
56	7522128	4905121F6	SNP00096086	155	295	C	C	T	P44	n/a	n/a	n/a	n/a
56	7522128	4905121F6	SNP00138849	195	335	G	G	T	E57	n/a	n/a	n/a	n/a
56	7522128	7716304J1	SNP00096086	217	274	C	C	T	T37	n/a	n/a	n/a	n/a
56	7522128	7716304J1	SNP00138849	257	314	G	G	T	G50	n/a	n/a	n/a	n/a
56	7522128	7738540J1	SNP00023420	479	713	A	A	G	G183	0.88	0.95	0.91	0.91
60	7513053	1217338R6	SNP00073109	98	1166	C	C	T	noncoding	n/a	n/a	n/a	n/a
60	7513053	1217338R6	SNP00107661	285	1354	T	T	C	noncoding	n/a	n/a	n/a	n/a
60	7513053	1281418H1	SNP00065340	220	838	C	C	T	L219	n/a	n/a	n/a	n/a
60	7513053	1349276H1	SNP00013133	32	765	T	C	T	R194	n/a	n/a	n/a	n/a
60	7513053	1923810H1	SNP00046483	10	5	T	T	G	noncoding	n/a	n/a	n/a	n/a
60	7513053	1996323R6	SNP00107661	60	1357	T	T	C	noncoding	n/a	n/a	n/a	n/a
60	7513053	2134965F6	SNP00137298	30	172	C	C	T	noncoding	n/a	n/a	n/a	n/a
60	7513053	4871044H1	SNP00046484	35	424	C	T	C	P81	n/a	n/a	n/a	n/a
60	7513053	603201T6	SNP00073109	129	1203	C	C	T	noncoding	n/a	n/a	n/a	n/a
60	7513053	7737790H1	SNP00065340	114	839	C	C	T	P219	n/a	n/a	n/a	n/a
63	7513718	1700077F6	SNP00053074	251	793	G	G	C	E236	n/a	n/a	n/a	n/a
64	7514003	137725F1	SNP00023895	240	1317	C	C	T	noncoding	n/a	n/a	n/a	n/a
64	7514003	1890540T6	SNP00023895	34	1466	C	C	T	noncoding	n/a	n/a	n/a	n/a
64	7514003	2861361T6	SNP00023895	141	1357	C	C	T	noncoding	n/a	n/a	n/a	n/a
64	7514003	7458847H1	SNP00023895	72	1319	C	C	T	noncoding	n/a	n/a	n/a	n/a